U.S. DEPARTMENT OF COMMERCE
Patent and Trademark Office

SEARCH REQUEST FORM

Requestor's Name: _____ Serial Number: _____

Date: _____ Phone: _____ Art Unit: _____

Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevent citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevent claim(s).

STAFF USE ONLY

Date completed: 11/19

Searcher: S. Hutzell 305-4205

Terminal time: 30 [illegible] 17

Elapsed time: 18

CPU time: _____

Total time: _____

Number of Searches: _____

Number of Databases: _____

Search Site

_____ STIC

__✓__ CM-1 CM-2

_____ Pre-S

Type of Search

__3__ N.A. Sequence

_____ A.A. Sequence

_____ Structure

_____ Bibliographic

Vendors

_____ IG

_____ STN

_____ Dialog

_____ APS

_____ Geninfo

_____ SDC

_____ DARC/Questel

__✓__ Other

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 16, 2002, 07:31:18 ; Search time 3268.87 Seconds
(without alignments)
17031.470 Million cell updates/sec

Title: M33262_COPY_1193_3105
Perfect score: 1913
Sequence: 1 GGTACCAGACGGCGTGAGGA.....TCTGGGGATGTCTCTAAATT 1913

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1913	100.0	9860	14	SIU89134	U89134 Simian-Huma
2	1913	100.0	10000	14	AF038398	AF038398 Simian-Hu
3	1913	100.0	10006	14	AF038399	AF038399 Simian-Hu
4	1913	100.0	10006	14	AF041850	AF041850 Simian-Hu
5	1913	100.0	10279	14	AY033146	AY033146 Simian im
6	1913	100.0	10279	14	AY033233	AY033233 Simian im
7	1913	100.0	10535	14	SIVMM239	M33262 Simian (mac
8	1909.8	99.8	9860	14	AF217181	AF217181 Simian-Hu
9	1884.8	98.5	10274	14	SIV1A11AA	M76764 Simian immu
10	1860	97.2	10277	14	SIVMM32H	D01065 Simian immu
11	1833.4	95.8	10277	14	SIVMM251	M19499 Simian (mac
12	1788.6	93.5	9646	14	RESIVMXX	Y00277 Simian immu
13	1772.6	92.7	9628	14	SIVMNE	M32741 Simian immu
14	1756.6	91.8	9664	14	SIU79412	U79412 Simian immu
15	1507.4	78.8	1533	14	SIVJ11P55	M74934 Simian immu
16	1505.8	78.7	1533	14	SIVJ12P55	M74935 Simian immu
17	1505.8	78.7	1533	14	SIVJ3P55	M74926 Simian immu
18	1505.8	78.7	1533	14	SIVJ8P55	M74931 Simian immu
19	1501	78.5	1533	14	SIVJ2P55	M74925 Simian immu
20	1495.4	78.2	1532	14	SIVJ7P55	M74930 Simian immu
21	1490.2	77.9	1532	6	AX180751	AX180751 Sequence
22	1471.8	76.9	1521	14	SIVJ10P55	M74933 Simian immu
23	1468.2	76.7	10241	14	RESIVSMM	X14307 Simian immu
24	1467.4	76.7	1533	14	SIVJ5P55	M74928 Simian immu
25	1458.2	76.2	1520	14	SIVJ6P55	M74929 Simian immu
26	1447.6	75.7	10184	14	AF077017	AF077017 Simian im
27	1444.6	75.5	10289	14	SIU72748	U72748 Simian immu
28	1424.6	74.5	9638	14	SIVGAGAB	M80194 Simian immu
29	1424.6	74.5	9675	14	SIVGAGAA	M80193 Simian immu
30	1423.4	74.4	9996	14	SIVSMMPBJA	M31325 Simian immu
31	1422.2	74.3	1521	14	SIU95965	U95965 Simian immu
32	1411.4	73.8	1533	6	AX180750	AX180750 Sequence
33	1407.6	73.6	10246	14	SIV6P12	L09211 Simian immu
34	1407.6	73.6	10246	14	SIV6P6	L09212 Simian immu
35	1407.6	73.6	10246	14	SIV6P9	L09213 Simian immu
36	1406	73.5	9996	14	SIVSMMPBJB'	L03295 Simian immu
37	1284.6	67.2	9892	14	SIVSTM	M83293 Simian (stu
38	1216	63.6	9599	14	AF208027	AF208027 HIV-2 iso
39	1179.4	61.7	9525	14	HIV2MDS	Z48731 Human immun
40	1174.4	61.4	10163	6	AR153663	AR153663 Sequence
41	1174.4	61.4	10172	14	HIV2U22047	U22047 Human immun
42	1170.8	61.2	10359	14	HIV2BEN	M30502 Human immun
43	1162.8	60.8	10312	14	HIV2U38293	U38293 Human immun
44	1158.6	60.6	9672	6	I09513	I09513 Sequence 1
45	1158.6	60.6	9672	14	HIV2ST	M31113 Human immun

ALIGNMENTS

RESULT 1
SIU89134
LOCUS SIU89134 9860 bp DNA linear VRL 21-MAY-1997
DEFINITION Simian-Human immunodeficiency virus strain SHIV-89.6P, complete genome.
ACCESSION U89134
VERSION U89134.1 GI:2108163
KEYWORDS .
SOURCE Simian-Human immunodeficiency virus.
ORGANISM Simian-Human immunodeficiency virus
Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate lentivirus group.
REFERENCE 1 (bases 1 to 9860)
AUTHORS Karlsson,G.B., Halloran,M., Li,J., Park,I.W., Gomila,R., Reimann,K.A., Axthelm,M.K., Iliff,S.A., Letvin,N.L. and Sodroski,J.

```
  TITLE     Characterization of molecularly cloned simian-human
            immunodeficiency viruses causing rapid CD4+ lymphocyte depletion in
            rhesus monkeys
  JOURNAL   J. Virol. 71 (6), 4218-4225 (1997)
  MEDLINE   97296224
   PUBMED   9151808
REFERENCE   2  (bases 1 to 9860)
  AUTHORS   Karlsson,G.B., Halloran,M., Li,J., Park,I., Gomila,R.,
            Reimann,K.A., Axthelm,M.K., Iliff,S.A., Letvin,N.L. and Sodroski,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-FEB-1997) Human Retrovirology, Dana-Farber Cancer
            Institute, 44 Binney Street, Boston, MA 02115, USA
FEATURES             Location/Qualifiers
     source          1. .9860
                     /organism="Simian-Human immunodeficiency virus"
                     /proviral
                     /strain="SHIV-89.6P"
                     /db_xref="taxon:57667"
                     /clone="SHIV-89.6KB9"
     source          1. .5946
                     /organism="Simian-Human immunodeficiency virus"
                     /db_xref="taxon:57667"
                     /note="derived from SIV"
     source          5947. .8778
                     /organism="Simian-Human immunodeficiency virus"
                     /db_xref="taxon:57667"
                     /note="derived from HIV-1"
     source          8779. .9860
                     /organism="Simian-Human immunodeficiency virus"
                     /db_xref="taxon:57667"
                     /note="derived from SIV"
     LTR             1. .301
                     /note="5' LTR"
     misc_feature    1. .177
                     /note="R region"
     misc_feature    178. .301
                     /note="U5 region"
     gene            536. .2068
                     /gene="gag"
     CDS             536. .2068
                     /gene="gag"
                     /codon_start=1
                     /product="gag polyprotein"
                     /protein_id="AAC57419.1"
                     /db_xref="GI:2108164"
                     /translation="MGVRNSVLSGKKADELEKIRLRPNGKKKYMLKHVVWAANELDRF
                     GLAESLLENKEGCQKILSVLAPLVPTGSENLKSLYNTVCVIWCIHAEEKVKHTEEAKQ
                     IVQRHLVVETGTTETMPKTSRPTAPSSGRGGNYPVQQIGGNYVHLPLSPRTLNAWVKL
                     IEEKKFGAEVVPGFQALSEGCTPYDINQMLNCVGDHQAAMQIIRDIINEEAADWDLQH
                     PQPAPQQGQLREPSGSDIAGTTSSVDEQIQWMYRQQNPIPVGNIYRRWIQLGLQKCVR
                     MYNPTNILDVKQGPKEPFQSYVDRFYKSLRAEQTDAAVKNWMTQTLLIQNANPDCKLV
                     LKGLGVNPTLEEMLTACQGVGGPGQKARLMAEALKEALAPVPIPFAAAQQRGPRKPIK
                     CWNCGKEGHSARQCRAPRRQGCWKCGKMDHVMAKCPDRQAGFLGLGPWGKKPRNFPMA
                     QVHQGLMPTAPPEDPAVDLLKNYMQLGKQQREKQRESREKPYKEVTEDLLHLNSLFGG
                     DQ"
     gene            <1711. .4893
                     /gene="pol"
     CDS             <1714. .4893
                     /gene="pol"
                     /codon_start=1
                     /product="pol polyprotein"
                     /protein_id="AAC57420.1"
                     /db_xref="GI:2108165"
                     /translation="LELWERGTLCKAMQSPKKTGMLEMWKNGPCYGQMPRQTGGFFRP
                     WSMGKEAPQFPHGSSASGADANCSPRGPSCGSAKELHAVGQAAERKAERKQREALQGG
                     DRGFAAPQFSLWRRPVVTAHIEGQPVEVLLDTGADDSIVTGIELGPHYTPKIVGGIGG
                     FINTKEYKNVEIEVLGKRIKGTIMTGDTPINIFGRNLLTALGMSLNFPIAKVEPVKVA
                     LKPGKDGPKLKQWPLSKEKIVALREICEKMEKDGQLEEAPPTNPYNTPTFAIKKKDKN
                     KWRMLIDFRELNRVTQDFTEVQLGIPHPAGLAKRKRITVLDIGDAYFSIPLDEEFRQY
                     TAFTLPSVNNAEPGKRYIYKVLPQGWKGSPAIFQYTMRHVLEPFRKANPDVTLVQYMD
                     DILIASDRTDLEHDRVVLQSKELLNSIGFSTPEEKFQKDPPFQWMGYELWPTKWKLQK
                     IELPQRETWTVNDIQKLVGVLNWAAQIYPGIKTKHLCRLIRGKMTLTEEVQWTEMAEA
                     EYEENKIILSQEQEGCYYQEGKPLEATVIKSQDNQWSYKIHQEDKILKVGKFAKIKNT
                     HTNGVRLLAHVIQKIGKEAIVIWGQVPKFHLPVEKDVWEQWWTDYWQVTWIPEWDFIS
                     TPPLVRLVFNLVKDPIEGEETYYTDGSCNKQSKEGKAGYITDRGKDKVKVLEQTTNQQ
                     AELEAFLMALTDSGPKANIIVDSQYVMGIITGCPTESESRLVNQIIEEMIKKSEIYVA
                     WVPAHKGIGGNQEIDHLVSQGIRQVLFLEKIEPAQEEHDKYHSNVKELVFKFGLPRIV
                     ARQIVDTCDKCHQKGEAIHGQANSDLGTWQMDCTHLEGKIIIVAVHVASGFIEAEVIP
                     QETGRQTALFLLKLAGRWPITHLHTDNGANFASQEVKMVAWWAGIEHTFGVPYNPQSQ
                     GVVEAMNHHLKNQIDRIREQANSVETIVLMAVHCMNFKRRGGIGDMTPAERLINMITT
                     EQEIQFQQSKNSKFKNFRVYYREGRDQLWKGPGELLWKGEGAVILKVGTDIKVVPRRK
                     AKIIKDYGGGKEVDSSSHMEDTGEAREVA"
     gene            4823. .5467
                     /gene="vif"
     CDS             4823. .5467
                     /gene="vif"
                     /codon_start=1
                     /product="vif protein"
                     /protein_id="AAC57421.1"
                     /db_xref="GI:2108166"
                     /translation="MEEEKRWIAVPTWRIPERLERWHSLIKYLKYKTKDLQKVCYVPH
                     FKVGWAWWTCSRVIFPLQEGSHLEVQGYWHLTPEKGWLSTYAVRITWYSKNFWTDVTP
                     NYADILLHSTYFPCFTAGEVRRAIRGEQLLSCCRFPRAHKYQVPSLQYLALKVVSDVR
                     SQGENPTWKQWRRDNRRGLRMAKQNSRGDKQRGGKPPTKGANFPGLAKVLGILA"
     gene            5295. .5633
                     /gene="vpx"
     CDS             5295. .5633
                     /gene="vpx"
                     /codon_start=1
                     /product="vpx protein"
                     /protein_id="AAC57422.1"
                     /db_xref="GI:2108167"
                     /translation="MSDPRERIPPGNSGEETIGEAFEWLNRTVEEINREAVNHLPREL
                     IFQVWQRSWEYWHDEQGMSPSYVKYRYLCLIQKALFMHCKKGCRCLGEGHGAGGWRPG
                     PPPPPPPGLA"
     gene            5634. .5939
                     /gene="vpr"
     CDS             5634. .5939
                     /gene="vpr"
                     /codon_start=1
                     /product="vpr protein"
                     /protein_id="AAC57423.1"
                     /db_xref="GI:2108168"
                     /translation="MEERPPENEGPQREPWDEWVVEVLEELKEEALKHFDPRLLTALG
                     NHIYNRHGDTLEGAGELIRILQRALFMHFRGGCIHSRIGQPGGGNPLSAIPPSRSML"
     gene            5948. .8580
                     /gene="tat"
     CDS             join(5948. .6162,8487. .8580)
                     /gene="tat"
                     /codon_start=1
                     /product="tat protein"
                     /protein_id="AAC57424.1"
                     /db_xref="GI:2108169"
                     /translation="MEPVDPRLEPWKHPGSKPKTACTNCYCKKCCFHCQVCFTTKALG
                     ISYGRKKRRQRRRAHQNSQTHQASLSKQPSSQPRGDPTGPKEQKKKVERETETDPVHQ
                     "
     gene            6087. .8761
                     /gene="rev"
     CDS             join(6087. .6162,8487. .8761)
                     /gene="rev"
                     /codon_start=1
                     /product="rev protein"
                     /protein_id="AAC57425.1"
                     /db_xref="GI:2108170"
                     /translation="MAGRSGDSDEELIRTVRLIKLLYQSNPPPSLEGTRQARRNRRRR
                     WRERQRQIRSISERILGTYLGRSAEPVPLQLPPLERLTLDCNEDCGTSGTQGVGSPQI
                     LVESPTVLESGTKE"
     gene            6179. .6424
                     /gene="vpu"
     CDS             6179. .6424
                     /gene="vpu"
                     /codon_start=1
                     /product="vpu protein"
                     /protein_id="AAC57426.1"
                     /db_xref="GI:2108171"
                     /translation="MQPIQIAIVALVVAIIIAIVVWSIVIIEYRKILRQRKIDRLIDR
```

```
                LIERAEDSGNESEGEISALVEMGVEMGHHAPWDVDDL"
    gene        6339. .8948
                /gene="env"
    CDS         6339. .8948
                /gene="env"
                /codon_start=1
                /product="env polyprotein"
                /protein_id="AAC57427.1"
                /db_xref="GI:2108172"
                /translation="MRVKEKYQHLWRWGWRWGTMLLGMLMICSATEKLWVTVYYGVPV
                WREATTTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEVVLGNVTENFNMWKNNMVD
                QMHEDIISLWDESLKPCVKLTPLCVTLNCTNLNITKNTTNLTSSSWGMMEEGEIKNCS
                FYITTSIRNKVKKEYALFNRLDVVPVKNTSNTKYRLISCNTSVITQACPKVSFQPIPI
                HYCVPAGFAILKCNNKTFNGSGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAEEDIVIR
                SEDFTDNVKTIIVQLNESVVINCTRPNNNTRERLSIGPGRAFYARRNIIGDIRQAHCN
                ISRAKWNNTLQQIVIKLREKFRNKTIAFNQSSGGDPEIVMHSFNCGGEFFYCNTAQLF
                NSTWNVAGGTNGTEGNDIITLQCRIKQIINMWQKVGKAMYAPPITGQIRCSSNITGLL
                LTRDGGNSTETETEIFRPGGGDMRDNWRSELYKYKVVRIEPIGVAPTRAKRRTVQREK
                RAVGIGAVFLGFLGAAGSTMGAASVTLTVQARLLLSGIVQQQNNLLRAIEAQQNMLRL
                TVWGIKQLQARVLALERYLRDQQLMGIWGCSGKLICTTSVPWNVSWSNKSVDDIWNNM
                TWMEWEREIDNYTDYIYDLLEKSQTQQEKNEKELLELDKWASLWNWFDITNWLWYIRL
                FIMIVGGLIGLRIVFAVLSIVNRVRQGYSPLSFQTLLPASRGPDRPEGTEEEGGERDR
                DRSGPSVNGSLALIWDDLRSLCLFSYHRLRDLLLIVTRIVELLGRRGWEALKYWWNLL
                QYWSQELKNSAVSLLQYGWSYFHEAVQAVWRSATETLAGAWGDLWETLRRGGRWILAI
                PRRIRQGLELTLL"
    gene        8782. .9573
                /gene="nef"
    CDS         8782. .9573
                /gene="nef"
                /codon_start=1
                /product="nef protein"
                /protein_id="AAC57428.1"
                /db_xref="GI:2108173"
                /translation="MGGAISMRRSRPSGDLRQRLLRARGETYGRLLGEVEDGYSQSPG
                GLDKGLSSLSCEGQKYNQGQYMNTPWRNPAEEREKLAYRKQNMDDIDEEDDDLVGVSV
                RPKVPLRTMSYKLAIDMSHFIKEKGGLEGIYYSARRHRILDIYLEKEEGIIPDWQDYT
                SGPGIRYPKTFGWLWKLVPVNVSDEAQEDEEHYLMHPAQTSQWDDPWGEVLAWKFDPT
                LAYTYEAYVRYPEEFGSKSGLSEEEVRRRLTARGLLNMADKKETR"
    LTR         9167. .9860
                /note="3' LTR"
    misc_feature 9167. .9683

  Query Match            100.0%;  Score 1913;  DB 14;  Length 9860;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1913;  Conservative    0;  Mismatches   0;  Indels    0;  Gaps    0;

Qy       1 GGTACCAGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGCAAC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     420 GGTACCAGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGCAAC 479

Qy      61 ACAAAAAAGAAATAGCTGTCTTTTATCCAGGAAGGGGTAATAAGATAGAGTGGGAGATGG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     480 ACAAAAAAGAAATAGCTGTCTTTTATCCAGGAAGGGGTAATAAGATAGAGTGGGAGATGG 539

Qy     121 GCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCTAC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     540 GCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCTAC 599

Qy     181 GACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTAG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     600 GACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTAG 659

Qy     241 ATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTTT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     660 ATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTTT 719

Qy     301 CGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACTG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     720 CGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACTG 779

Qy     361 TCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAAC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     780 TCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAAC 839

Qy     421 AGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAAACAA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     840 AGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAAACAA 899

Qy     481 GTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGGTG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     900 GTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGGTG 959

Qy     541 GTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGATAG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     960 GTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGATAG 1019

Qy     601 AGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGTTGCA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1020 AGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGTTGCA 1079

Qy     661 CCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCAGA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1080 CCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCAGA 1139

Qy     721 TTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACCAG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1140 TTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACCAG 1199

Qy     781 CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACTAGTT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1200 CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACTAGTT 1259

Qy     841 CAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGCAACA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1260 CAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGCAACA 1319

Qy     901 TTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCAACAA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1320 TTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCAACAA 1379

Qy     961 ACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1380 ACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCT 1439

Qy    1021 ACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACAC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1440 ACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACAC 1499

Qy    1081 TGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATC 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1500 TGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATC 1559

Qy    1141 CCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCTA 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1560 CCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCTA 1619

Qy    1201 GATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCAGCAG 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1620 GATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCAGCAG 1679

Qy    1261 CCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGACACT 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1680 CCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGACACT 1739

Qy    1321 CTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATGGACC 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1740 CTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATGGACC 1799

Qy    1381 ATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGGGGAA 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1800 ATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGGGGAA 1859

Qy    1441 AGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGCTCCCC 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1860 AGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGCTCCCC 1919
```

```
Qy   1501 CAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAGAGAG 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1920 CAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAGAGAG 1979

Qy   1561 AAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCACC 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1980 AAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCACC 2039

Qy   1621 TCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAGA 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2040 TCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAGA 2099

Qy   1681 AGTATTACTGGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCACA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2100 AGTATTACTGGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCACA 2159

Qy   1741 TTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAAA 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2160 TTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAAA 2219

Qy   1801 TGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCCC 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2220 TGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCCC 2279

Qy   1861 GATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAATT 1913
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2280 GATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAATT 2332
```

RESULT 2
AF038398

```
LOCUS       AF038398               10000 bp    DNA     linear   VRL 02-FEB-1998
DEFINITION  Simian-Human immunodeficiency virus strain SHIV-89.6, complete
            genome.
ACCESSION   AF038398
VERSION     AF038398.1  GI:2828036
KEYWORDS    .
SOURCE      Simian-Human immunodeficiency virus.
  ORGANISM  Simian-Human immunodeficiency virus
            Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
            lentivirus group.
REFERENCE   1  (bases 1 to 10000)
  AUTHORS   Reimann,K.A., Li,J.T., Voss,G., Lekutis,C., Tenner-Racz,K.,
            Racz,P., Lin,W., Montefiori,D.C., Lee-Parritz,D.E., Lu,Y.,
            Collman,R.G., Sodroski,J. and Letvin,N.L.
  TITLE     An env gene derived from a primary human immunodeficiency virus
            type 1 isolate confers high in vivo replicative capacity to a
            chimeric simian/human immunodeficiency virus in rhesus monkeys
  JOURNAL   J. Virol. 70 (5), 3198-3206 (1996)
  MEDLINE   96186751
   PUBMED   8627800
REFERENCE   2  (bases 1 to 10000)
  AUTHORS   Reimann,K.A., Li,J.T., Voss,G., Lekutis,C., Tenner-Racz,K.,
            Racz,P., Lin,W., Montefiori,D.C., Lee-Parritz,D.E., Lu,Y.,
            Collman,R.G., Sodroski,J. and Letvin,N.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-DEC-1997) Human Retrovirology, Dana-Farber Cancer
            Institute, 44 Binney Street, Boston, MA 02115, USA
FEATURES             Location/Qualifiers
     source          1. .10000
                     /organism="Simian-Human immunodeficiency virus"
                     /proviral
                     /strain="SHIV-89.6"
                     /db_xref="taxon:57667"
     LTR             1. .301
                     /note="5' long terminal repeat"
     repeat_region   1. .177
                     /note="R region"
     misc_feature    178. .301
                     /note="U5 region"
     gene            536. .2068
                     /gene="gag"
     CDS             536. .2068
                     /gene="gag"
                     /codon_start=1
                     /product="gag"
                     /protein_id="AAB99958.1"
                     /db_xref="GI:2828037"
                     /translation="MGVRNSVLSGKKADELEKIRLRPNGKKKYMLKHVVWAANELDRF
                     GLAESLLENKEGCQKILSVLAPLVPTGSENLKSLYNTVCVIWCIHAEEKVKHTEEAKQ
                     IVQRHLVVETGTTETMPKTSRPTAPSSGRGGNYPVQQIGGNYVHLPLSPRTLNAWVKL
                     IEEKKFGAEVVPGFQALSEGCTPYDINQMLNCVGDHQAAMQIIRDIINEEAADWDLQH
                     PQPAPQQGQLREPSGSDIAGTTSSVDEQIQWMYRQQNPIPVGNIYRRWIQLGLQKCVR
                     MYNPTNILDVKQGPKEPFQSYVDRFYKSLRAEQTDAAVKNWMTQTLLIQNANPDCKLV
                     LKGLGVNPTLEEMLTACQGVGGPGQKARLMAEALKEALAPVPIPFAAAQQRGPRKPIK
                     CWNCGKEGHSARQCRAPRRQGCWKCGKMDHVMAKCPDRQAGFLGLGPWGKKPRNFPMA
                     QVHQGLMPTAPPEDPAVDLLKNYMQLGKQQREKQRESREKPYKEVTEDLLHLNSLFGG
                     DQ"
     gene            <1714. .4893
                     /gene="pol"
     CDS             <1714. .4893
                     /gene="pol"
                     /codon_start=1
                     /product="pol"
                     /protein_id="AAB99959.1"
                     /db_xref="GI:2828038"
                     /translation="LELWERGTLCKAMQSPKKTGMLEMWKNGPCYGQMPRQTGGFFRP
                     WSMGKEAPQFPHGSSASGADANCSPRGPSCGSAKELHAVGQAAERKAERKQREALQGG
                     DRGFAAPQFSLWRRPVVTAHIEGQPVEVLLDTGADDSIVTGIELGPHYTPKIVGGIGG
                     FINTKEYKNVEIEVLGKRIKGTIMTGDTPINIFGRNLLTALGMSLNFPIAKVEPVKVA
                     LKPGKDGPKLKQWPLSKEKIVALREICEKMEKDGQLEEAPPTNPYNTPTFAIKKKDKN
                     KWRMLIDFRELNRVTQDFTEVQLGIPHPAGLAKRKRITVLDIGDAYFSIPLDEEFRQY
                     TAFTLPSVNNAEPGKRYIYKVLPQGWKGSPAIFQYTMRHVLEPFRKANPDVTLVQYMD
                     DILIASDRTDLEHDRVVLQSKELLNSIGFSTPEEKFQKDPPFQWMGYELWPTKWKLQK
                     IELPQRETWTVNDIQKLVGVLNWAAQIYPGIKTKHLCRLIRGKMTLTEEVQWTEMAEA
                     EYEENKIILSQEQEGCYYQEGKPLEATVIKSQDNQWSYKIHQEDKILKVGKFAKIKNT
                     HTNGVRLLAHVIQKIGKEAIVIWGQVPKFHLPVEKDVWEQWWTDYWQVTWIPEWDFIS
                     TPPLVRLVFNLVKDPIEGEETYYTDGSCNKQSKEGKAGYITDRGKDKVKVLEQTTNQQ
                     AELEAFLMALTDSGPKANIIVDSQYVMGIITGCPTESESRLVNQIIEEMIKKSEIYVA
                     WVPAHKGIGGNQEIDHLVSQGIRQVLFLEKIEPAQEEHDKYHSNVKELVFKFGLPRIV
                     ARQIVDTCDKCHQKGEAIHGQANSDLGTWQMDCTHLEGKIIIVAVHVASGFIEAEVIP
                     QETGRQTALFLLKLAGRWPITHLHTDNGANFASQEVKMVAWWAGIEHTFGVPYNPQSQ
                     GVVEAMNHHLKNQIDRIREQANSVETIVLMAVHCMNFKRRGGIGDMTPAERLINMITT
                     EQEIQFQQSKNSKFKNFRVYYREGRDQLWKGPGELLWKGEGAVILKVGTDIKVVPRRK
                     AKIIKDYGGGKEVDSSSHMEDTGEAREVA"
     gene            4823. .5467
                     /gene="vif"
     CDS             4823. .5467
                     /gene="vif"
                     /codon_start=1
                     /product="vif"
                     /protein_id="AAB99960.1"
                     /db_xref="GI:2828039"
                     /translation="MEEEKRWIAVPTWRIPERLERWHSLIKYLKYKTKDLQKVCYVPH
                     FKVGWAWWTCSRVIFPLQEGSHLEVQGYWHLTPEKGWLSTYAVRITWYSKNFWTDVTP
                     NYADILLHSTYFPCFTAGEVRRAIRGEQLLSCCRFPRAHKYQVPSLQYLALKVVSDVR
                     SQGENPTWKQWRRDNRRGLRMAKQNSRGDKQRGGKPPTKGANFPGLAKVLGILA"
     gene            5295. .5633
                     /gene="vpx"
     CDS             5295. .5633
                     /gene="vpx"
                     /codon_start=1
                     /product="vpx"
                     /protein_id="AAB99961.1"
                     /db_xref="GI:2828040"
                     /translation="MSDPRERIPPGNSGEETIGEAFEWLNRTVEEINREAVNHLPREL
                     IFQVWQRSWEYWHDEQGMSPSYVKYRYLCLIQKALFMHCKKGCRCLGEGHGAGGWRPG
                     PPPPPPPGLA"
     gene            5634. .5939
                     /gene="vpr"
     CDS             5634. .5939
                     /gene="vpr"
                     /codon_start=1
                     /product="vpr"
                     /protein_id="AAB99962.1"
                     /db_xref="GI:2828041"
```

```
                /translation="MEERPPENEGPQREPWDEWVVEVLEELKEEALKHFDPRLLTALG
                NHIYNRHGDTLEGAGELIRILQRALFMHFRGGCIHSRIGQPGGGNPLSAIPPSRSML"
gene            5948. .8761
                /gene="tat"
CDS             join(5948. .6162,8487. .8577)
                /gene="tat"
                /codon_start=1
                /product="tat"
                /protein_id="AAB99963.1"
                /db_xref="GI:2828042"
                /translation="MEPVDPRLEPWKHPGSKPKTACTNCYCKKCCFHCQVCFITKALG
                ISYGRKKRRQRRRAHQNSQTHQASLSKQPSSQPRGDPTGPKEQKKKVERETETDPVH"
gene            6087. .8761
                /gene="rev"
CDS             join(6087. .6162,8487. .8761)
                /gene="rev"
                /codon_start=1
                /product="rev"
                /protein_id="AAB99964.1"
                /db_xref="GI:2828043"
                /translation="MAGRSGDSDEELIRTVRLIKLLYQSNPPPSLEGTRQARRNRRRR
                WRERQRQIRSISERILGTYLGRSAEPVPLQLPPLERLTLDCNEDCGTSGTQGVGSPQI
                LVESPTVLESGTKE"
gene            6179. .6424
                /gene="vpu"
CDS             6179. .6424
                /gene="vpu"
                /codon_start=1
                /product="vpu"
                /protein_id="AAB99965.1"
                /db_xref="GI:2828044"
                /translation="MQPIQIAIVALVVAIIIAIVVWSIVIIEYRKILRQRKIDRLIDR
                LIERAEDSGNESEGEISALVEMGVEMGHHAPWDVDDL"
gene            6339. .8903
                /gene="env"
CDS             6339. .8903
                /gene="env"
                /codon_start=1
                /product="env"
                /protein_id="AAB99966.1"
                /db_xref="GI:2828045"
                /translation="MRVKEKYQHLWRWGWRWGTMLLGMLMICSATEKLWVTVYYGVPV
                WREATTTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEVVLGNVTENFNMWKNNMVD
                QMHEDIISLWDESLKPCVKLTPLCVTLNCTNLNITKNTTNPTSSSWGMMEKGEIKNCS
                FYITTSIRNKVKKEYALFNRLDVVPIENTNNTKYRLISCNTSVITQACPKVSFQPIPI
                HYCVPAGFAMLKCNNKTFNGSGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAEEDIVIR
                SENFTDNAKTIIVQLNESVVINCTRPNNNTRRRLSIGPGRAFYARRNIIGDIRQAHCN
                ISRAKWNNTLQQIVIKLREKFRNKTIAFNQSSGGDPEIVMHSFNCGGEFFYCNTAQLF
                NSTWNVTGGTNGTEGNDIITLQCRIKQIINMWQKVGKAMYAPPITGQIRCSSNITGLL
                LTRDGGNSTETETEIFRPGGGDMRDNWRSELYKYKVVRIEPIGVAPTRAKRRTVQREK
                RAVGIGAVFLGFLGAAGSTMGAASVTLTVQARLLLSGIVQQQNNLLRAIEAQQHMLQL
                TVWGIKQLQARVLALERYLRDQQLMGIWGCSGKLICTTSVPWNVSWSNKSVDDIWNNM
                TWMEWEREIDNYTDYIYDLLEKSQTQQEKNEKELLELDKWASLWNWFDITNWLWYIRL
                FIMIVGGLIGLRIVFAVLSIVNRVRQGYSPLSFQTLLPASRGPDRPEGTEEEGGERDR
                DRSGPLVNGSLALIWDDLRSLCLFSYHRLRDLLLIVTRIVELLGRRGWEALKYWWNLL
                QYWSQELKNSAVSLLNATAIAVAEGTDRVIEVVQGACRAIRHIPRRIRQGLERILL"
gene            8922. .9713
                /gene="nef"
CDS             8922. .9713
                /gene="nef"
                /codon_start=1
                /product="nef protein"
                /protein_id="AAB99967.1"
                /db_xref="GI:2828046"
                /translation="MGGAISMRRSRPSGDLRQRLLRARGETYGRLLGEVEDGYSQSPG
                GLDKGLSSLSCEGQKYNQGQYMNTPWRNPAEEREKLAYRKQNMDDIDEEDDDLVGVSV
                RPKVPLRTMSYKLAIDMSHFIKEKGGLEGIYYSARRHRILDIYLEKEEGIIPDWQDYT
                SGPGIRYPKTFGWLWKLVPVNVSDEAQEDEEHYLMHPAQTSQWDDPWGEVLAWKFDPT
                LAYTYEAYVRYPEEFGSKSGLSEEEVRRRLTARGLLNMADKKETR"
LTR             9307. .10000
                /note="3' long terminal repeat"
misc_feature    9307. .9823
                /note="U3 region"
```

```
    repeat_region   9824. .10000
                    /note="R region"
BASE COUNT     3466 a   1817 c   2500 g   2217 t
ORIGIN

  Query Match             100.0%;  Score 1913;  DB 14;  Length 10000;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1913;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGTACCAGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGCAAC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    420 GGTACCAGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGCAAC 479

Qy     61 ACAAAAAAGAAATAGCTGTCTTTTATCCAGGAAGGGGTAATAAGATAGAGTGGGAGATGG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    480 ACAAAAAAGAAATAGCTGTCTTTTATCCAGGAAGGGGTAATAAGATAGAGTGGGAGATGG 539

Qy    121 GCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCTAC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    540 GCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCTAC 599

Qy    181 GACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTAG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    600 GACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTAG 659

Qy    241 ATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTTT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    660 ATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTTT 719

Qy    301 CGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACTG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    720 CGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACTG 779

Qy    361 TCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAAC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    780 TCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAAC 839

Qy    421 AGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAAACAA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    840 AGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAAACAA 899

Qy    481 GTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGGTG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    900 GTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGGTG 959

Qy    541 GTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGATAG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    960 GTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGATAG 1019

Qy    601 AGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGTTGCA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1020 AGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGTTGCA 1079

Qy    661 CCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCAGA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1080 CCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCAGA 1139

Qy    721 TTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACCAG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1140 TTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACCAG 1199

Qy    781 CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACTAGTT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1200 CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACTAGTT 1259

Qy    841 CAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGCAACA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1260 CAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGCAACA 1319

Qy    901 TTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCAACAA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1320 TTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCAACAA 1379
```

```
Qy     961 ACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCT 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1380 ACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCT 1439

Qy    1021 ACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACAC 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1440 ACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACAC 1499

Qy    1081 TGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATC 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1500 TGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATC 1559

Qy    1141 CCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCTA 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1560 CCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCTA 1619

Qy    1201 GATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCAGCAG 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1620 GATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCAGCAG 1679

Qy    1261 CCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGACACT 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1680 CCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGACACT 1739

Qy    1321 CTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATGGACC 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1740 CTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATGGACC 1799

Qy    1381 ATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGGGGAA 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1800 ATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGGGGAA 1859

Qy    1441 AGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGCTCCCC 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1860 AGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGCTCCCC 1919

Qy    1501 CAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAGAGAG 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1920 CAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAGAGAG 1979

Qy    1561 AAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCACC 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1980 AAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCACC 2039

Qy    1621 TCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAGA 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2040 TCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAGA 2099

Qy    1681 AGTATTACTGGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCACA 1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2100 AGTATTACTGGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCACA 2159

Qy    1741 TTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAAA 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2160 TTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAAA 2219

Qy    1801 TGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCCC 1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2220 TGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCCC 2279

Qy    1861 GATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAATT 1913
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2280 GATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAATT 2332
```

```
RESULT 3
AF038399
LOCUS       AF038399               10006 bp    DNA     linear   VRL 02-FEB-1998
DEFINITION  Simian-Human immunodeficiency virus strain SHIV-4, complete genome.
ACCESSION   AF038399
VERSION     AF038399.1  GI:2828047
KEYWORDS    .
SOURCE      Simian-Human immunodeficiency virus.
  ORGANISM  Simian-Human immunodeficiency virus
            Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
            lentivirus group.
REFERENCE   1  (bases 1 to 10006)
  AUTHORS   Li,J., Lord,C.I., Haseltine,W., Letvin,N.L. and Sodroski,J.
  TITLE     Infection of cynomolgus monkeys with a chimeric HIV-1/SIVmac virus
            that expresses the HIV-1 envelope glycoproteins
  JOURNAL   J. Acquir. Immune Defic. Syndr. 5 (7), 639-646 (1992)
  MEDLINE   92309177
   PUBMED   1613662
REFERENCE   2  (bases 1 to 10006)
  AUTHORS   Reimann,K.A., Li,J.T., Voss,G., Lekutis,C., Tenner-Racz,K.,
            Racz,P., Lin,W., Montefiori,D.C., Lee-Parritz,D.E., Lu,Y.,
            Collman,R.G., Sodroski,J. and Letvin,N.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-DEC-1997) Human Retrovirology, Dana-Farber Cancer
            Institute, 44 Binney Street, Boston, MA 02115, USA
FEATURES             Location/Qualifiers
     source          1. .10006
                     /organism="Simian-Human immunodeficiency virus"
                     /proviral
                     /strain="SHIV-4; HXBc2"
                     /db_xref="taxon:57667"
     LTR             1. .301
                     /note="5' long terminal repeat"
     repeat_region   1. .177
                     /note="R region"
     misc_feature    178. .301
                     /note="U5 region"
     gene            536. .2068
                     /gene="gag"
     CDS             536. .2068
                     /gene="gag"
                     /codon_start=1
                     /product="gag"
                     /protein_id="AAB99968.1"
                     /db_xref="GI:2828048"
                     /translation="MGVRNSVLSGKKADELEKIRLRPNGKKKYMLKHVVWAANELDRF
                     GLAESLLENKEGCQKILSVLAPLVPTGSENLKSLYNTVCVIWCIHAEEKVKHTEEAKQ
                     IVQRHLVVETGTTETMPKTSRPTAPSSGRGGNYPVQQIGGNYVHLPLSPRTLNAWVKL
                     IEEKKFGAEVVPGFQALSEGCTPYDINQMLNCVGDHQAAMQIIRDIINEEAADWDLQH
                     PQPAPQQGQLREPSGSDIAGTTSSVDEQIQWMYRQQNPIPVGNIYRRWIQLGLQKCVR
                     MYNPTNILDVKQGPKEPFQSYVDRFYKSLRAEQTDAAVKNWMTQTLLIQNANPDCKLV
                     LKGLGVNPTLEEMLTACQGVGGPGQKARLMAEALKEALAPVPIPFAAAQQRGPRKPIK
                     CWNCGKEGHSARQCRAPRRQGCWKCGKMDHVMAKCPDRQAGFLGLGPWGKKPRNFPMA
                     QVHQGLMPTAPPEDPAVDLLKNYMQLGKQQREKQRESREKPYKEVTEDLLHLNSLFGG
                     DQ"
     gene            <1714. .4893
                     /gene="pol"
     CDS             <1714. .4893
                     /gene="pol"
                     /codon_start=1
                     /product="pol"
                     /protein_id="AAB99969.1"
                     /db_xref="GI:2828049"
                     /translation="LELWERGTLCKAMQSPKKTGMLEMWKNGPCYGQMPRQTGGFFRP
                     WSMGKEAPQFPHGSSASGADANCSPRGPSCGSAKELHAVGQAAERKAERKQREALQGG
                     DRGFAAPQFSLWRRPVVTAHIEGQPVEVLLDTGADDSIVTGIELGPHYTPKIVGGIGG
                     FINTKEYKNVEIEVLGKRIKGTIMTGDTPINIFGRNLLTALGMSLNFPIAKVEPVKVA
                     LKPGKDGPKLKQWPLSKEKIVALREICEKMEKDGQLEEAPPTNPYNTPTFAIKKKDKN
                     KWRMLIDFRELNRVTQDFTEVQLGIPHPAGLAKRKRITVLDIGDAYFSIPLDEEFRQY
                     TAFTLPSVNNAEPGKRYIYKVLPQGWKGSPAIFQYTMRHVLEPFRKANPDVTLVQYMD
                     DILIASDRTDLEHDRVVLQSKELLNSIGFSTPEEKFQKDPPFQWMGYELWPTKWKLQK
                     IELPQRETWTVNDIQKLVGVLNWAAQIYPGIKTKHLCRLIRGKMTLTEEVQWTEMAEA
                     EYEENKIILSQEQEGCYYQEGKPLEATVIKSQDNQWSYKIHQEDKILKVGKFAKIKNT
                     HTNGVRLLAHVIQKIGKEAIVIWGQVPKFHLPVEKDVWEQWWTDYWQVTWIPEWDFIS
                     TPPLVRLVFNLVKDPIEGEETYYTDGSCNKQSKEGKAGYITDRGKDKVKVLEQTTNQQ
                     AELEAFLMALTDSGPKANIIVDSQYVMGIITGCPTESESRLVNQIIEEMIKKSEIYVA
                     WVPAHKGIGGNQEIDHLVSQGIRQVLFLEKIEPAQEEHDKYHSNVKELVFKFGLPRIV
                     ARQIVDTCDKCHQKGEAIHGQANSDLGTWQMDCTHLEGKIIIVAVHVASGFIEAEVIP
                     QETGRQTALFLLKLAGRWPITHLHTDNGANFASQEVKMVAWWAGIEHTFGVPYNPQSQ
                     GVVEAMNHHLKNQIDRIREQANSVETIVLMAVHCMNFKRRGGIGDMTPAERLINMITT
                     EQEIQFQQSKNSKFKNFRVYYREGRDQLWKGPGELLWKGEGAVILKVGTDIKVVPRRK
```

```
                AKIIKDYGGGKEVDSSSHMEDTGEAREVA"
gene            4823. .5467
                /gene="vif"
CDS             4823. .5467
                /gene="vif"
                /codon_start=1
                /product="vif"
                /protein_id="AAB99970.1"
                /db_xref="GI:2828050"
                /translation="MEEEKRWIAVPTWRIPERLERWHSLIKYLKYKTKDLQKVCYVPH
                FKVGWAWWTCSRVIFPLQEGSHLEVQGYWHLTPEKGWLSTYAVRITWYSKNFWTDVTP
                NYADILLHSTYFPCFTAGEVRRAIRGEQLLSCCRFPRAHKYQVPSLQYLALKVVSDVR
                SQGENPTWKQWRRDNRRGLRMAKQNSRGDKQRGGKPPTKGANFPGLAKVLGILA"
gene            5295. .5633
                /gene="vpx"
CDS             5295. .5633
                /gene="vpx"
                /codon_start=1
                /product="vpx"
                /protein_id="AAB99971.1"
                /db_xref="GI:2828051"
                /translation="MSDPRERIPPGNSGEETIGEAFEWLNRTVEEINREAVNHLPREL
                IFQVWQRSWEYWHDEQGMSPSYVKYRYLCLIQKALFMHCKKGCRCLGEGHGAGGWRPG
                PPPPPPPGLA"
gene            5634. .5939
                /gene="vpr"
CDS             5634. .5939
                /gene="vpr"
                /codon_start=1
                /product="vpr"
                /protein_id="AAB99972.1"
                /db_xref="GI:2828052"
                /translation="MEERPPENEGPQREPWDEWVVEVLEELKEEALKHFDPRLLTALG
                NHIYNRHGDTLEGAGELIRILQRALFMHFRGGCIHSRIGQPGGGNPLSAIPPSRSML"
gene            5948. .8580
                /gene="tat"
CDS             join(5948. .6162,8493. .8538)
                /gene="tat"
                /codon_start=1
                /product="tat"
                /protein_id="AAB99973.1"
                /db_xref="GI:2828053"
                /translation="MEPVDPRLEPWKHPGSKPKTACTNCYCKKCCFHCQVCFITKALG
                ISYGRKKRRQRRRAHQNSQTHQASLSKQPTSQPRGDPTGPKE"
gene            6087. .8767
                /gene="rev"
CDS             join(6087. .6162,8493. .8767)
                /gene="rev"
                /codon_start=1
                /product="rev"
                /protein_id="AAB99974.1"
                /db_xref="GI:2828054"
                /translation="MAGRSGDSDEELIRTVRLIKLLYQSNPPPNPEGTRQARRNRRRR
                WRERQRQIHSISERILGTYLGRSAEPVPLQLPPLERLTLDCNEDCGTSGTQGVGSPQI
                LVESPTVLESGTKE"
gene            6179. .6424
                /gene="vpu"
CDS             6179. .6424
                /gene="vpu"
                /codon_start=1
                /product="vpu"
                /protein_id="AAB99975.1"
                /db_xref="GI:2828055"
                /translation="MQPIQIAIVALVVAIIIAIVVWSIVIIEYRKILRQRKIDRLIDR
                LIERAEDSGNESEGEISALVEMGVEMGHHAPWDVDDL"
gene            6339. .8909
                /gene="env"
CDS             6339. .8909
                /gene="env"
                /codon_start=1
                /product="env"
                /protein_id="AAB99976.1"
                /db_xref="GI:2828056"
                /translation="MRVKEKYQHLWRWGWRWGTMLLGMLMICSATEKLWVTVYYGVPV
                WKEATTTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEVVLVNVTENFNMWKNDMVE
                QMHEDIISLWDQSLKPCVKLTPLCVSLKCTDLKNDTNTNSSSGRMIMEKGEIKNCSFN
                ISTSIRGKVQKEYAFFYKLDIIPIDNDTTSYKLTSCNTSVITQACPKVSFEPIPIHYC
                APAGFAILKCNNKTFNGTGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAEEEVVIRSVN
                FTDNAKTIIVQLNTSVEINCTRPNNNTRKRIRIQRGPGRAFVTIGKIGNMRQAHCNIS
                RAKWNNTLKQIASKLREQFGNNKTIIFKQSSGGDPEIVTHSFNCGGEFFYCNSTQLFN
                STWFNSTWSTEGSNNTEGSDTITLPCRIKQIINMWQKVGKAMYAPPISGQIRCSSNIT
                GLLLTRDGGNSNNESEIFRPGGGDMRDNWRSELYKYKVVKIEPLGVAPTKAKRRVVQR
                EKRAVGIGALFLGFLGAAGSTMGAASMTLTVQARQLLSGIVQQQNNLLRAIEAQQHLL
                QLTVWGIKQLQARILAVERYLKDQQLLGIWGCSGKLICTTAVPWNASWSNKSLEQIWN
                HTTWMEWDREINNYTSLIHSLIEESQNQQEKNEQELLELDKWASLWNWFNITNWLWYI
                KLFIMIVGGLVGLRIVFAVLSIVNRVRQGYSPLSFQTHLPTPRGPDRPEGIEEEGGER
                DRDRSIRLVNGSLALIWDDLRSLCLFSYHRLRDLLLIVTRIVELLGRRGWEALKYWWN
                LLQYWSQELKNSAVSLLNATAIAVAEGTDRVIEVVQGACRAIRHIPRRIRQGLERILL
                "
gene            8928. .9719
                /gene="nef"
CDS             8928. .9719
                /gene="nef"
                /codon_start=1
                /product="nef protein"
                /protein_id="AAB99977.1"
                /db_xref="GI:2828057"
                /translation="MGGAISMRRSRPSGDLRQRLLRARGETYGRLLGEVEDGYSQSPG
                GLDKGLSSLSCEGQKYNQGQYMNTPWRNPAEEREKLAYRKQNMDDIDEEDDDLVGVSV
                RPKVPLRTMSYKLAIDMSHFIKEKGGLEGIYYSARRHRILDIYLEKEEGIIPDWQDYT
                SGPGIRYPKTFGWLWKLVPVNVSDEAQEDEEHYLMHPAQTSQWDDPWGEVLAWKFDPT
                LAYTYEAYVRYPEEFGSKSGLSEEEVRRRLTARGLLNMADKKETR"
LTR             9313. .10006
                /note="3' long terminal repeat"
misc_feature    9313. .9829
                /note="U3 region"
repeat_region   9830. .10006
                /note="R region"
BASE COUNT     3468 a   1827 c   2499 g   2212 t
ORIGIN
```

```
  Query Match             100.0%;  Score 1913;  DB 14;  Length 10006;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1913;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GGTACCAGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGCAAC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     420 GGTACCAGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGCAAC 479

Qy      61 ACAAAAAAGAAATAGCTGTCTTTTATCCAGGAAGGGGTAATAAGATAGAGTGGGAGATGG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     480 ACAAAAAAGAAATAGCTGTCTTTTATCCAGGAAGGGGTAATAAGATAGAGTGGGAGATGG 539

Qy     121 GCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCTAC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     540 GCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCTAC 599

Qy     181 GACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTAG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     600 GACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTAG 659

Qy     241 ATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTTT 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     660 ATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTTT 719

Qy     301 CGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACTG 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     720 CGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACTG 779

Qy     361 TCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAAC 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     780 TCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAAC 839

Qy     421 AGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAAACAA 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     840 AGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAAACAA 899
```

```
Qy    481 GTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGGTG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    900 GTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGGTG 959

Qy    541 GTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGATAG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    960 GTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGATAG 1019

Qy    601 AGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGTTGCA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1020 AGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGTTGCA 1079

Qy    661 CCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCAGA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1080 CCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCAGA 1139

Qy    721 TTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACCAG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1140 TTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACCAG 1199

Qy    781 CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACTAGTT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1200 CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACTAGTT 1259

Qy    841 CAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGCAACA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1260 CAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGCAACA 1319

Qy    901 TTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCAACAA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1320 TTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCAACAA 1379

Qy    961 ACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1380 ACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCT 1439

Qy   1021 ACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACAC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1440 ACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACAC 1499

Qy   1081 TGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATC 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1500 TGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATC 1559

Qy   1141 CCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCTA 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1560 CCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCTA 1619

Qy   1201 GATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCAGCAG 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1620 GATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCAGCAG 1679

Qy   1261 CCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGACACT 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1680 CCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGACACT 1739

Qy   1321 CTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATGGACC 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1740 CTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATGGACC 1799

Qy   1381 ATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGGGGAA 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1800 ATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGGGGAA 1859

Qy   1441 AGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGCTCCCC 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1860 AGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGCTCCCC 1919

Qy   1501 CAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAGAGAG 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1920 CAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAGAGAG 1979
```

```
Qy   1561 AAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCACC 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1980 AAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCACC 2039

Qy   1621 TCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAGA 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2040 TCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAGA 2099

Qy   1681 AGTATTACTGGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCACA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2100 AGTATTACTGGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCACA 2159

Qy   1741 TTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAAA 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2160 TTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAAA 2219

Qy   1801 TGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCCC 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2220 TGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCCC 2279

Qy   1861 GATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAATT 1913
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2280 GATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAATT 2332
```

RESULT 4
AF041850

```
LOCUS       AF041850               10006 bp    DNA     linear   VRL 03-FEB-1999
DEFINITION  Simian-Human immunodeficiency virus clone SHIV-HXBc2P 3.2, complete
            genome.
ACCESSION   AF041850
VERSION     AF041850.1  GI:2895564
KEYWORDS    .
SOURCE      Simian-Human immunodeficiency virus.
  ORGANISM  Simian-Human immunodeficiency virus
            Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
            lentivirus group.
REFERENCE   1  (bases 1 to 10006)
  AUTHORS   Cayabyab,M., Karlsson,G.B., Etemad-Moghadam,B.A., Hofmann,W.,
            Steenbeke,T., Halloran,M., Fanton,J.W., Axthelm,M.K., Letvin,N.L.
            and Sodroski,J.G.
  TITLE     Changes in human immunodeficiency virus type 1 envelope
            glycoproteins responsible for the pathogenicity of a multiply
            passaged simian-human immunodeficiency virus (SHIV-HXBc2)
  JOURNAL   J. Virol. 73 (2), 976-984 (1999)
  MEDLINE   99098984
   PUBMED   9882298
REFERENCE   2  (bases 1 to 10006)
  AUTHORS   Cayabyab,M., Karlsson,G.B., Etemad-Moghadam,B., Hofmann,W.,
            Halloran,M., Axthelm,M.W., Letvin,N.L. and Sodroski,J.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JAN-1998) Human Retrovirology, Dana-Farber Cancer
            Institute, 44 Binney Street, Boston, MA 02115, USA
FEATURES             Location/Qualifiers
     source          1. .10006
                     /organism="Simian-Human immunodeficiency virus"
                     /proviral
                     /db_xref="taxon:57667"
                     /clone="SHIV-HXBc2P 3.2"
                     /note="SHIV-HXBc2P3.2 is a molecularly cloned virus from
                     the pathogenic stock KU-1; Ku-1 was generated by multiply
                     passaging SHIV-4 (GenBank Accession Number AF038399)"
     LTR             1. .301
                     /note="5'LTR"
     repeat_region   1. .177
                     /note="R-region"
     repeat_region   178. .301
                     /note="U5-region"
     gene            536. .2068
                     /gene="gag"
     CDS             536. .2068
                     /gene="gag"
                     /codon_start=1
```

```
                /product="gag protein"
                /protein_id="AAD12134.1"
                /db_xref="GI:2895565"
                /translation="MGVRNSVLSGKKADELEKIRLRPNGKKKYMLKHVVWAANELDRF
                GLAESLLENKEGCQKILSVLAPLVPTGSENLKSLYNTVCVIWCIHAEEKVKHTEEAKQ
                IVQRHLVVETGTTETMPKTSRPTAPSSGRGGNYPVQQIGGNYVHLPLSPRTLNAWVKL
                IEEKKFGAEVVPGFQALSEGCTPYDINQMLNCVGDHQAAMQIIRDIINEEAADWDLQH
                PQPAPQQGQLREPSGSDIAGTTSSVDEQIQWMYRQQNPIPVGNIYRRWIQLGLQKCVR
                MYNPTNILDVKQGPKEPFQSYVDRFYKSLRAEQTDAAVKNWMTQTLLIQNANPDCKLV
                LKGLGVNPTLEEMLTACQGVGGPGQKARLMAEALKEALAPVPIPFAAAQQRGPRKPIK
                CWNCGKEGHSARQCRAPRRQGCWKCGKMDHVMAKCPDRQAGFLGLGPWGKKPRNFPMA
                QVHQGLMPTAPPEDPAVDLLKNYMQLGKQQREKQRESREKPYKEVTEDLLHLNSLFGG
                DQ"
gene            <1714. .4893
                /gene="pol"
CDS             <1714. .4893
                /gene="pol"
                /codon_start=1
                /product="pol protein"
                /protein_id="AAD12135.1"
                /db_xref="GI:2895566"
                /translation="LELWERGTLCKAMQSPKKTGMLEMWKNGPCYGQMPRQTGGFFRP
                WSMGKEAPQFPHGSSASGADANCSPRGPSCGSAKELHAVGQAAERKAERKQREALQGG
                DRGFAAPQFSLWRRPVVTAHIEGQPVEVLLDTGADDSIVTGIELGPHYTPKIVGGIGG
                FINTKEYKNVEIEVLGKRIKGTIMTGDTPINIFGRNLLTALGMSLNFPIAKVEPVKVA
                LKPGKDGPKLKQWPLSKEKIVALREICEKMEKDGQLEEAPPTNPYNTPTFAIKKKDKN
                KWRMLIDFRELNRVTQDFTEVQLGIPHPAGLAKRKRITVLDIGDAYFSIPLDEEFRQY
                TAFTLPSVNNAEPGKRYIYKVLPQGWKGSPAIFQYTMRHVLEPFRKANPDVTLVQYMD
                DILIASDRTDLEHDRVVLQSKELLNSIGFSTPEEKFQKDPPFQWMGYELWPTKWKLQK
                IELPQRETWTVNDIQKLVGVLNWAAQIYPGIKTKHLCRLIRGKMTLTEEVQWTEMAEA
                EYEENKIILSQEQEGCYYQEGKPLEATVIKSQDNQWSYKIHQEDKILKVGKFAKIKNT
                HTNGVRLLAHVIQKIGKEAIVIWGQVPKFHLPVEKDVWEQWWTDYWQVTWIPEWDFIS
                TPPLVRLVFNLVKDPIEGEETYYTDGSCNKQSKEGKAGYITDRGKDKVKVLEQTTNQQ
                AELEAFLMALTDSGPKANIIVDSQYVMGIITGCPTESESRLVNQIIEEMIKKSEIYVA
                WVPAHKGIGGNQEIDHLVSQGIRQVLFLEKIEPAQEEHDKYHSNVKELVFKFGLPRIV
                ARQIVDTCDKCHQKGEAIHGQANSDLGTWQMDCTHLEGKIIIVAVHVASGFIEAEVIP
                QETGRQTALFLLKLAGRWPITHLHTDNGANFASQEVKMVAWWAGIEHTFGVPYNPQSQ
                GVVEAMNHHLKNQIDRIREQANSVETIVLMAVHCMNFKRRGGIGDMTPAERLINMITT
                EQEIQFQQSKNSKFKNFRVYYREGRDQLWKGPGELLWKGEGAVILKVGTDIKVVPRRK
                AKIIKDYGGGKEVDSSSHMEDTGEAREVA"
gene            4823. .5467
                /gene="vif"
CDS             4823. .5467
                /gene="vif"
                /codon_start=1
                /product="vif protein"
                /protein_id="AAD12136.1"
                /db_xref="GI:2895567"
                /translation="MEEEKRWIAVPTWRIPERLERWHSLIKYLKYKTKDLQKVCYVPH
                FKVGWAWWTCSRVIFPLQEGSHLEVQGYWHLTPEKGWLSTYAVRITWYSKNFWTDVTP
                NYADILLHSTYFPCFTAGEVRRAIRGEQLLSCCRFPRAHKYQVPSLQYLALKVVSDVR
                SQGENPTWKQWRRDNRRGLRMAKQNSRGDKQRGGKPPTKGANFPGLAKVLGILA"
gene            5295. .5633
                /gene="vpx"
CDS             5295. .5633
                /gene="vpx"
                /codon_start=1
                /product="vpx protein"
                /protein_id="AAD12137.1"
                /db_xref="GI:2895568"
                /translation="MSDPRERIPPGNSGEETIGEAFEWLNRTVEEINREAVNHLPREL
                IFQVWQRSWEYWHDEQGMSPSYVKYRYLCLIQKALFMHCKKGCRCLGEGHGAGGWRPG
                PPPPPPPGLA"
gene            5634. .5939
                /gene="vpr"
CDS             5634. .5939
                /gene="vpr"
                /codon_start=1
                /product="vpr protein"
                /protein_id="AAD12138.1"
                /db_xref="GI:2895569"
                /translation="MEERPPENEGPQREPWDEWVVEVLEELKEEALKHFDPRLLTALG
                NHIYNRHGDTLEGAGELIRILQRALFMHFRGGCIHSRIGQPGGGNPLSAIPPSRSML"

gene            5948. .8538
                /gene="tat"
CDS             join(5948. .6162,8493. .8538)
                /gene="tat"
                /codon_start=1
                /product="tat protein"
                /protein_id="AAD12139.1"
                /db_xref="GI:2895570"
                /translation="MEPVDPRLEPWKHPGSKPKTACTNCYCKKCCFHCQVCFITKALG
                ISYGRKKRRQRRRAHQNSQTHQASLSKQPTSQPRGDPTGPKE"
gene            6087. .8767
                /gene="rev"
CDS             join(6087. .6162,8493. .8767)
                /gene="rev"
                /codon_start=1
                /product="rev protein"
                /protein_id="AAD12140.1"
                /db_xref="GI:2895571"
                /translation="MAGRSGDSDEELIRTVRLIKLLYQSNPPPNPEGTRQARRNRRRR
                WRERQRQIHSISERILGTYLGRSAEPVPLQLPPLERLTLDCNEDCGTSGTQGVGSPQI
                LVESPTVLESGTKE"
gene            6179. .6424
                /gene="vpu"
CDS             6179. .6424
                /gene="vpu"
                /codon_start=1
                /product="vpu protein"
                /protein_id="AAD12141.1"
                /db_xref="GI:2895572"
                /translation="MQPIQIAIVALVVAIIIAIVVWSIVIIEYRKILRQRKIDRLIDR
                LIERAEDSGNESEGEISALVEMGVEMGHHAPWDVDDL"
gene            6339. .8909
                /gene="env"
CDS             6339. .8909
                /gene="env"
                /codon_start=1
                /product="envelope glycoprotein"
                /protein_id="AAD12142.1"
                /db_xref="GI:2895573"
                /translation="MRVKEKYQHLWRWGWRWGTMLLGMLMICSATEKLWVTVYYGVPV
                WKEATTTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEVVLVNVTENFNMWKNDMVE
                QMHEDIISLWDQSLKPCVKLTPLCVSLNCTDLKNDTNTNSSSGRMIMEKGEIKNCSFN
                ISTSIRGKVQKEYAFFYKLDIIPIDNDTTSYTLTSCNTSVISQACPKVSFEPIPIHYC
                APAGFAILKCNNKTFNGTGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAEEEVVIRSVN
                FMDNAKTIIVQLNTSVEINCTRPSNNTIKRIRIQRGPGRAFVTMGKIGDMRQAHCNIS
                RAKWNNTLKQIASKLREQFGNNKTIIFKQSSGGDPEIVTHSFNCGGEFFYCNSTQLFN
                STWFNSTWSTEGSNNTEGSDTITLPCRIKQIINMWQKVGKAMYAPPISGQIRCSSNIT
                GLLLTRDGGKGNNESEIFRPGGGDMRDNWRSELYKYKVVKIEPLGVAPTKAKRRVVQR
                EKRAVGIGALFLGFLGAAGSTMGAASMTLTVQARQLLSGIVQQQNNLLRAIEAQQHLL
                QLTVWGIKQLQARILAVERYLKDQQLLGIWGCSGKLICTTAVPWNASWSNKSLEQIWN
                HMTWMEWDREINNYTSLIHSLIEESQNQQEKNEQELLELDKWASLWNWFDITNWLWYI
                KLFIMIVGGLVGLRIVFAVLSIVNRVRQGYSPLSFQTHLPTPRGPDRPEGIEEEGGER
                DRDRSIRLVNGSLALIWDDLRSLCLFSYHRLRDLLLIVTRIVELLGRRGWEALKYWWN
                LLQYWSQELKNSAVSLLNATAIAVAEGTDRVIEVVQGACRAIRHIPRRIRQGLERILL
                "
gene            8928. .9719
                /gene="nef"
CDS             8928. .9719
                /gene="nef"
                /codon_start=1
                /product="nef protein"
                /protein_id="AAD12143.1"
                /db_xref="GI:2895574"
                /translation="MGGAISMRRSRPSGDLRQRLLRARGETYGRLLGEVEDGYSQSPG
                GLDKGLSSLSCEGQKYNQGQYMNTPWRNPAEEREKLAYRKQNMDDIDEEDDDLVGVSV
                RPKVPLRTMSYKLAIDMSHFIKEKGGLEGIYYSARRHRILDIYLEKEEGIIPDWQDYT
                SGPGIRYPKTFGWLWKLVPVNVSDEAQEDEEHYLMHPAQTSQWDDPWGEVLAWKFDPT
                LAYTYEAYVRYPEEFGSKSGLSEEEVRRRLTARGLLNMADKKETR"
LTR             9313. .10006
                /note="5'LTR"
repeat_region   9313. .9829
                /note="U3-region"
repeat_region   9830. .10006
```

```
                     /note="R-region"
BASE COUNT     3459 a   1824 c   2505 g   2218 t
ORIGIN

  Query Match           100.0%;  Score 1913;  DB 14;  Length 10006;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1913;  Conservative    0;  Mismatches   0;  Indels    0;  Gaps    0;

Qy       1 GGTACCAGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGCAAC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     420 GGTACCAGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGCAAC 479

Qy      61 ACAAAAAAGAAATAGCTGTCTTTTATCCAGGAAGGGGTAATAAGATAGAGTGGGAGATGG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     480 ACAAAAAAGAAATAGCTGTCTTTTATCCAGGAAGGGGTAATAAGATAGAGTGGGAGATGG 539

Qy     121 GCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCTAC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     540 GCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCTAC 599

Qy     181 GACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTAG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     600 GACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTAG 659

Qy     241 ATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTTT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     660 ATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTTT 719

Qy     301 CGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACTG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     720 CGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACTG 779

Qy     361 TCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAAC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     780 TCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAAC 839

Qy     421 AGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAAACAA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     840 AGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAAACAA 899

Qy     481 GTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGGTG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     900 GTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGGTG 959

Qy     541 GTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGATAG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     960 GTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGATAG 1019

Qy     601 AGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGTTGCA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1020 AGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGTTGCA 1079

Qy     661 CCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCAGA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1080 CCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCAGA 1139

Qy     721 TTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACCAG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1140 TTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACCAG 1199

Qy     781 CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACTAGTT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1200 CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACTAGTT 1259

Qy     841 CAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGCAACA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1260 CAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGCAACA 1319

Qy     901 TTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCAACAA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1320 TTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCAACAA 1379

Qy     961 ACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1380 ACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCT 1439

Qy    1021 ACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACAC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1440 ACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACAC 1499

Qy    1081 TGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATC 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1500 TGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATC 1559

Qy    1141 CCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCTA 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1560 CCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCTA 1619

Qy    1201 GATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCAGCAG 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1620 GATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCAGCAG 1679

Qy    1261 CCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGACACT 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1680 CCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGACACT 1739

Qy    1321 CTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATGGACC 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1740 CTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATGGACC 1799

Qy    1381 ATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGGGGAA 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1800 ATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGGGGAA 1859

Qy    1441 AGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGCTCCCC 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1860 AGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGCTCCCC 1919

Qy    1501 CAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAGAGAG 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1920 CAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAGAGAG 1979

Qy    1561 AAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCACC 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1980 AAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCACC 2039

Qy    1621 TCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAGA 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2040 TCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAGA 2099

Qy    1681 AGTATTACTGGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCACA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2100 AGTATTACTGGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCACA 2159

Qy    1741 TTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAAA 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2160 TTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAAA 2219

Qy    1801 TGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCCC 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2220 TGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCCC 2279

Qy    1861 GATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAATT 1913
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2280 GATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAATT 2332


RESULT 5
AY033146
LOCUS       AY033146               10279 bp    DNA     linear   VRL 05-JUL-2001
DEFINITION  Simian immunodeficiency virus clone 17E-Fr, complete sequence.
ACCESSION   AY033146
VERSION     AY033146.1  GI:14600336
KEYWORDS    .
SOURCE      Simian immunodeficiency virus 17E-Fr.
  ORGANISM  Simian immunodeficiency virus 17E-Fr
```

```
          Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
          lentivirus group.
REFERENCE 1  (bases 1 to 10279)
  AUTHORS Flaherty,M.T., Hauer,D.A., Mankowski,J.L., Zink,M.C. and
          Clements,J.E.
  TITLE   Molecular and biological characterization of a neurovirulent
          molecular clone of simian immunodeficiency virus
  JOURNAL J. Virol. 71 (8), 5790-5798 (1997)
  MEDLINE 97366637
   PUBMED 9223467
REFERENCE 2  (bases 1 to 10279)
  AUTHORS Flaherty,M.T., Hauer,D.A., Mankowski,J.L., Zink,M.C. and
          Clements,J.E.
  TITLE   Direct Submission
  JOURNAL Submitted (24-APR-2001) Comp. Med., Johns Hopkins U. Sch. of Med.,
          600 N. Wolfe St., Baltimore, MD 21287, USA
FEATURES             Location/Qualifiers
     source          1. .10279
                     /organism="Simian immunodeficiency virus 17E-Fr"
                     /proviral
                     /db_xref="taxon:160753"
                     /clone="17E-Fr"
                     /note="recombinant of SIV/17E-Br and SIVmac239; envelope,
                     nef, and 3' LTR from SIV/17E-Br swarm were cloned in place
                     of the same genes in an infectious molecular clone of
                     SIVmac239"
     LTR             1. .818
                     /note="5' long terminal repeat"
     misc_feature    1. .517
                     /note="U3 region"
     misc_feature    518. .694
                     /note="R region"
     misc_feature    695. .818
                     /note="U5 region"
     gene            1053. .2585
                     /gene="gag"
     CDS             1053. .2585
                     /gene="gag"
                     /codon_start=1
                     /product="gag polyprotein"
                     /protein_id="AAK64193.1"
                     /db_xref="GI:14600340"
                     /translation="MGVRNSVLSGKKADELEKIRLRPNGKKKYMLKHVVWAANELDRF
                     GLAESLLENKEGCQKILSVLAPLVPTGSENLKSLYNTVCVIWCIHAEEKVKHTEEAKQ
                     IVQRHLVVETGTTETMPKTSRPTAPSSGRGGNYPVQQIGGNYVHLPLSPRTLNAWVKL
                     IEEKKFGAEVVPGFQALSEGCTPYDINQMLNCVGDHQAAMQIIRDIINEEAADWDLQH
                     PQPAPQQGQLREPSGSDIAGTTSSVDEQIQWMYRQQNPIPVGNIYRRWIQLGLQKCVR
                     MYNPTNILDVKQGPKEPFQSYVDRFYKSLRAEQTDAAVKNWMTQTLLIQNANPDCKLV
                     LKGLGVNPTLEEMLTACQGVGGPGQKARLMAEALKEALAPVPIPFAAAQQRGPRKPIK
                     CWNCGKEGHSARQCRAPRRQGCWKCGKMDHVMAKCPDRQAGFLGLGPWGKKPRNFPMA
                     QVHQGLMPTAPPEDPAVDLLKNYMQLGKQQREKQRESREKPYKEVTEDLLHLNSLFGG
                     DQ"
     gene            <2228. .5410
                     /gene="pol"
     CDS             <2228. .5410
                     /gene="pol"
                     /codon_start=1
                     /product="pol polyprotein"
                     /protein_id="AAK64194.1"
                     /db_xref="GI:14600341"
                     /translation="VLELWERGTLCKAMQSPKKTGMLEMWKNGPCYGQMPRQTGGFFR
                     PWSMGKEAPQFPHGSSASGADANCSPRGPSCGSAKELHAVGQAAERKAERKQREALQG
                     GDRGFAAPQFSLWRRPVVTAHIEGQPVEVLLDTGADDSIVTGIELGPHYTPKIVGGIG
                     GFINTKEYKNVEIEVLGKRIKGTIMTGDTPINIFGRNLLTALGMSLNFPIAKVEPVKV
                     ALKPGKDGPKLKQWPLSKEKIVALREICEKMEKDGQLEEAPPTNPYNTPTFAIKKKDK
                     NKWRMLIDFRELNRVTQDFTEVQLGIPHPAGLAKRKRITVLDIGDAYFSIPLDEEFRQ
                     YTAFTLPSVNNAEPGKRYIYKVLPQGWKGSPAIFQYTMRHVLEPFRKANPDVTLVQYM
                     DDILIASDRTDLEHDRVVLQSKELLNSIGFSTPEEKFQKDPPFQWMGYELWPTKWKLQ
                     KIELPQRETWTVNDIQKLVGVLNWAAQIYPGIKTKHLCRLIRGKMTLTEEVQWTEMAE
                     AEYEENKIILSQEQEGCYYQEGKPLEATVIKSQDNQWSYKIHQEDKILKVGKFAKIKN
                     THTNGVRLLAHVIQKIGKEAIVIWGQVPKFHLPVEKDVWEQWWTDYWQVTWIPEWDFI
                     STPPLVRLVFNLVKDPIEGEETYYTDGSCNKQSKEGKAGYITDRGKDKVKVLEQTTNQ
                     QAELEAFLMALTDSGPKANIIVDSQYVMGIITGCPTESESRLVNQIIEEMIKKSEIYV
                     AWVPAHKGIGGNQEIDHLVSQGIRQVLFLEKIEPAQEEHDKYHSNVKELVFKFGLPRI
                     VARQIVDTCDKCHQKGEAIHGQANSDLGTWQMDCTHLEGKIIIVAVHVASGFIEAEVI
                     PQETGRQTALFLLKLAGRWPITHLHTDNGANFASQEVKMVAWWAGIEHTFGVPYNPQS
                     QGVVEAMNHHLKNQIDRIREQANSVETIVLMAVHCMNFKRRGGIGDMTPAERLINMIT
                     TEQEIQFQQSKNSKFKNFRVYYREGRDQLWKGPGELLWKGEGAVILKVGTDIKVVPRR
                     KAKIIKDYGGGKEVDSSSHMEDTGEAREVA"
misc_feature         5084. .5099
                     /gene="pol"
                     /note="polypurine tract"
gene                 5340. .5984
                     /gene="vif"
CDS                  5340. .5984
                     /gene="vif"
                     /codon_start=1
                     /product="vif protein"
                     /protein_id="AAK64195.1"
                     /db_xref="GI:14600342"
                     /translation="MEEEKRWIAVPTWRIPERLERWHSLIKYLKYKTKDLQKVCYVPH
                     FKVGWAWWTCSRVIFPLQEGSHLEVQGYWHLTPEKGWLSTYAVRITWYSKNFWTDVTP
                     NYADILLHSTYFPCFTAGEVRRAIRGEQLLSCCRFPRAHKYQVPSLQYLALKVVSDVR
                     SQGENPTWKQWRRDNRRGLRMAKQNSRGDKQRGGKPPTKGANFPGLAKVLGILA"
gene                 5812. .6150
                     /gene="vpx"
CDS                  5812. .6150
                     /gene="vpx"
                     /codon_start=1
                     /product="vpx protein"
                     /protein_id="AAK64196.1"
                     /db_xref="GI:14600343"
                     /translation="MSDPRERIPPGNSGEETIGEAFEWLNRTVEEINREAVNHLPREL
                     IFQVWQRSWEYWHDEQGMSPSYVKYRYLCLIQKALFMHCKKGCRCLGEGHGAGGWRPG
                     PPPPPPPGLA"
gene                 6151. .6456
                     /gene="vpr"
CDS                  6151. .6456
                     /gene="vpr"
                     /codon_start=1
                     /product="vpr protein"
                     /protein_id="AAK64197.1"
                     /db_xref="GI:14600344"
                     /translation="MEERPPENEGPQREPWDEWVVEVLEELKEEALKHFDPRLLTALG
                     NHIYNRHGDTLEGAGELIRILQRALFMHFRGGCIHSRIGQPGGGNPLSAIPPSRSML"
gene                 6302. .8902
                     /gene="tat"
CDS                  join(6302. .6597,8803. .8902)
                     /gene="tat"
                     /note="transactivator of viral transcription"
                     /codon_start=1
                     /product="tat protein"
                     /protein_id="AAK64190.1"
                     /db_xref="GI:14600337"
                     /translation="METPLREQENSLESSNERSSCISEADASTPESANLGEEILSQLY
                     RPLEACYNTCYCKKCCYHCQFCFLKKGLGICYEQSRKRRRTPKKAKANTSSASNNRPI
                     SNRTRHCQPEKAKKETVEKAVATAPGLDR"
gene                 6528. .9059
                     /gene="rev"
CDS                  join(6528. .6597,8803. .9059)
                     /gene="rev"
                     /codon_start=1
                     /product="rev protein"
                     /protein_id="AAK64191.1"
                     /db_xref="GI:14600338"
                     /translation="MNNHEREEELRKRLRLIHLLHQTTDPYPTGPGTANQRRQRRRRW
                     RRRWQQLLALTDRIYSFPDPPTDTPLDLAIQQLQNLAIESIPDPPTNTPEALCDPTED
                     SRSPQD"
gene                 6604. .9243
                     /gene="env"
                     /note="contains premature stop codon"
CDS                  6604. .8895
                     /gene="env"
                     /codon_start=1
                     /product="truncated envelope glycoprotein"
                     /protein_id="AAK64192.1"
```

```
                     /db_xref="GI:14600339"
                     /translation="MGCLGNQLLIAILLLSVYGIYCTLYVTVFYGVPVWRNATIPLFC
                     ATKNRDTWGTTQCLPDNGDYSEMALNVTESFDAWNNTVTEQAIEDVWQLFETSIKPCV
                     KLSPLCITMRCNKSETDRWGLTKSITTTASTTSTTASARVDMVNETSSCIAQDNCAGL
                     EQEQMISCKFNMTGLNRDKKKEYNETWYSADLVCEQGNNTGNESRCYMNHCNTSVIKE
                     SCDKHYWDAIRFRYCAPPGYALLRCNDTNYSGFMPKCSKVVVSSCTRMMETQTSTWFG
                     FNGTRAENRTYIYWHGRDNRTIISLNKYYNLTIKCRRPGNKTVLPVTIMSGLVFHSQR
                     INDRPRQAWCWFGGKWKDAIKEVKQTIVKHPRYTGTNNTDKINLTAPRGGDPEVTFMW
                     TNCRGEFLYCKMNWFLNWVEDRNTANQKPKEQHKRNYVPCHIRQIINTWHKVGKNVYL
                     PPREGDLTCNSTVTSLIANIDWIDGNQTNITMSAEVAELYRLELGDYKLVEITPIGLA
                     PTDVKRYTTGGTSRNKRGVFVLGFLGFLATAGSAMGAASLTLTAQSRTLLAGIVQQQQ
                     QLLDVVKRQQELLRLTVWGTKNLQTRVTAIEKYLKDQAQLNTWGCAFRQVCHTTVPWP
                     NASLIPKWNNETWQEWERKVDFLEENITALLEEAQIQQEKNMYELQKLNSWNVFGNWF
                     DLASWIKYIQYGVYIVVGVILLRIVIYIVQVLAKLRQGYRPVFSSPPSYFQQTHIQQD
                     PALPTREGKEGDGGEGGGNSSWP"
     gene            9077. .9868
                     /gene="nef"
     CDS             9077. .9868
                     /gene="nef"
                     /codon_start=1
                     /product="nef protein"
                     /protein_id="AAK64198.1"
                     /db_xref="GI:14600345"
                     /translation="MGGAISMRRSRSSGDLRQRLLRARGETYGRLLGEVEDGYSQSPG
                     GLDKGLSSLSCEGQKYNQGQYMNTPWRNPAKEREKLAYRKQNMDDIDEQDDDLVGVSV
                     RPKVPLRTMSYKLAIDISHFIKEKGGLEGIYYSARRHRILDIYLEKEKGIIPDWQDYT
                     SGPGIRFPKTFGWLWKLVPVNVSDEAQEDEEHYLMHPAQTSQWDDPWGEVLAWKFDPT
                     LAYTYEAYVRYPEEFGSKSGLSEEEVRRRLTARGLLNMADKKETR"
     misc_feature    9445. .9460
                     /gene="nef"
                     /note="polypurine tract"
     LTR             9462. .10279
                     /note="3' long terminal repeat"
     misc_feature    9462. .9978
                     /note="U3 region"
     misc_feature    9979. .10155
                     /note="R region"
     polyA_signal    10132. .10137
     misc_feature    10156. .10279
                     /note="U5 region"
BASE COUNT     3470 a   1932 c   2566 g   2311 t
ORIGIN
```

Query Match 100.0%; Score 1913; DB 14; Length 10279;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1913; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 GGTACCAGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGCAAC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    937 GGTACCAGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGCAAC 996

Qy     61 ACAAAAAAGAAATAGCTGTCTTTTATCCAGGAAGGGGTAATAAGATAGAGTGGGAGATGG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    997 ACAAAAAAGAAATAGCTGTCTTTTATCCAGGAAGGGGTAATAAGATAGAGTGGGAGATGG 1056

Qy    121 GCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCTAC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1057 GCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCTAC 1116

Qy    181 GACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTAG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1117 GACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTAG 1176

Qy    241 ATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTTT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1177 ATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTTT 1236

Qy    301 CGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACTG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1237 CGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACTG 1296

Qy    361 TCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAAC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1297 TCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAAC 1356

Qy    421 AGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAAACAA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1357 AGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAAACAA 1416

Qy    481 GTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGGTG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1417 GTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGGTG 1476

Qy    541 GTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGATAG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1477 GTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGATAG 1536

Qy    601 AGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGTTGCA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1537 AGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGTTGCA 1596

Qy    661 CCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCAGA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1597 CCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCAGA 1656

Qy    721 TTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACCAG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1657 TTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACCAG 1716

Qy    781 CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACTAGTT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1717 CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACTAGTT 1776

Qy    841 CAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGCAACA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1777 CAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGCAACA 1836

Qy    901 TTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCAACAA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1837 TTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCAACAA 1896

Qy    961 ACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1897 ACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCT 1956

Qy   1021 ACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACAC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1957 ACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACAC 2016

Qy   1081 TGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATC 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2017 TGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATC 2076

Qy   1141 CCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCTA 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2077 CCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCTA 2136

Qy   1201 GATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCAGCAG 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2137 GATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCAGCAG 2196

Qy   1261 CCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGACACT 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2197 CCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGACACT 2256

Qy   1321 CTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATGGACC 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2257 CTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATGGACC 2316

Qy   1381 ATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGGGGAA 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2317 ATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGGGGAA 2376

Qy   1441 AGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGCTCCCC 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2377 AGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGCTCCCC 2436
```

```
Qy   1501 CAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAGAGAG 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2437 CAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAGAGAG 2496

Qy   1561 AAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCACC 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2497 AAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCACC 2556

Qy   1621 TCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAGA 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2557 TCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAGA 2616

Qy   1681 AGTATTACTGGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCACA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2617 AGTATTACTGGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCACA 2676

Qy   1741 TTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAAA 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2677 TTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAAA 2736

Qy   1801 TGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCCC 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2737 TGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCCC 2796

Qy   1861 GATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAATT 1913
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2797 GATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAATT 2849
```

RESULT 6
AY033233
LOCUS AY033233 10279 bp DNA linear VRL 05-JUL-2001
DEFINITION Simian immunodeficiency virus clone 17E-Cl, complete sequence.
ACCESSION AY033233
VERSION AY033233.1 GI:14600346
KEYWORDS .
SOURCE Simian immunodeficiency virus 17E-Cl.
ORGANISM Simian immunodeficiency virus 17E-Cl
Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate lentivirus group.
REFERENCE 1 (bases 1 to 10279)
AUTHORS Anderson,M.G., Hauer,D., Sharma,D.P., Joag,S.V., Narayan,O., Zink,M.C. and Clements,J.E.
TITLE Analysis of envelope changes acquired by SIVmac239 during neuroadaption in rhesus macaques
JOURNAL Virology 195 (2), 616-626 (1993)
MEDLINE 93331720
PUBMED 8337835
REFERENCE 2 (bases 1 to 10279)
AUTHORS Anderson,M.G., Hauer,D.A., Sharma,D.P., Joag,S.V., Narayan,O., Zink,M.C. and Clements,J.E.
TITLE Direct Submission
JOURNAL Submitted (24-APR-2001) Comp. Med., Johns Hopkins U. Sch. of Med., 600 N. Wolfe St., Baltimore, MD 21287, USA
FEATURES Location/Qualifiers
source 1. .10279
/organism="Simian immunodeficiency virus 17E-Cl"
/proviral
/db_xref="taxon:160754"
/clone="17E-Cl"
/note="recombinant of SIV/17E-Br and SIVmac239; SU portion of envelope from SIV/17E-Br swarm was cloned in place of the same gene in an infectious molecular clone of SIVmac239"
LTR 1. .818
/note="5' long terminal repeat"
misc_feature 1. .517
/note="U3 region"
misc_feature 518. .694
/note="R region"
misc_feature 695. .818
/note="U5 region"
gene 1053. .2585
/gene="gag"
CDS 1053. .2585
/gene="gag"
/codon_start=1
/product="gag polyprotein"
/protein_id="AAK64202.1"
/db_xref="GI:14600350"
/translation="MGVRNSVLSGKKADELEKIRLRPNGKKKYMLKHVVWAANELDRF
GLAESLLENKEGCQKILSVLAPLVPTGSENLKSLYNTVCVIWCIHAEEKVKHTEEAKQ
IVQRHLVVETGTTETMPKTSRPTAPSSGRGGNYPVQQIGGNYVHLPLSPRTLNAWVKL
IEEKKFGAEVVPGFQALSEGCTPYDINQMLNCVGDHQAAMQIIRDIINEEAADWDLQH
PQPAPQQGQLREPSGSDIAGTTSSVDEQIQWMYRQQNPIPVGNIYRRWIQLGLQKCVR
MYNPTNILDVKQGPKEPFQSYVDRFYKSLRAEQTDAAVKNWMTQTLLIQNANPDCKLV
LKGLGVNPTLEEMLTACQGVGGPGQKARLMAEALKEALAPVPIPFAAAQQRGPRKPIK
CWNCGKEGHSARQCRAPRRQGCWKCGKMDHVMAKCPDRQAGFLGLGPWGKKPRNFPMA
QVHQGLMPTAPPEDPAVDLLKNYMQLGKQQREKQRESREKPYKEVTEDLLHLNSLFGG
DQ"
gene <2228. .5410
/gene="pol"
CDS <2228. .5410
/gene="pol"
/codon_start=1
/product="pol polyprotein"
/protein_id="AAK64203.1"
/db_xref="GI:14600351"
/translation="VLELWERGTLCKAMQSPKKTGMLEMWKNGPCYGQMPRQTGGFFR
PWSMGKEAPQFPHGSSASGADANCSPRGPSCGSAKELHAVGQAAERKAERKQREALQG
GDRGFAAPQFSLWRRPVVTAHIEGQPVEVLLDTGADDSIVTGIELGPHYTPKIVGGIG
GFINTKEYKNVEIEVLGKRIKGTIMTGDTPINIFGRNLLTALGMSLNFPIAKVEPVKV
ALKPGKDGPKLKQWPLSKEKIVALREICEKMEKDGQLEEAPPTNPYNTPTFAIKKKDK
NKWRMLIDFRELNRVTQDFTEVQLGIPHPAGLAKRKRITVLDIGDAYFSIPLDEEFRQ
YTAFTLPSVNNAEPGKRYIYKVLPQGWKGSPAIFQYTMRHVLEPFRKANPDVTLVQYM
DDILIASDRTDLEHDRVVLQSKELLNSIGFSTPEEKFQKDPPFQWMGYELWPTKWKLQ
KIELPQRETWTVNDIQKLVGVLNWAAQIYPGIKTKHLCRLIRGKMTLTEEVQWTEMAE
AEYEENKIILSQEQEGCYYQEGKPLEATVIKSQDNQWSYKIHQEDKILKVGKFAKIKN
THTNGVRLLAHVIQKIGKEAIVIWGQVPKFHLPVEKDVWEQWWTDYWQVTWIPEWDFI
STPPLVRLVFNLVKDPIEGEETYYTDGSCNKQSKEGKAGYITDRGKDKVKVLEQTTNQ
QAELEAFLMALTDSGPKANIIVDSQYVMGIITGCPTESESRLVNQIIEEMIKKSEIYV
AWVPAHKGIGGNQEIDHLVSQGIRQVLFLEKIEPAQEEHDKYHSNVKELVFKFGLPRI
VARQIVDTCDKCHQKGEAIHGQANSDLGTWQMDCTHLEGKIIIVAVHVASGFIEAEVI
PQETGRQTALFLLKLAGRWPITHLHTDNGANFASQEVKMVAWWAGIEHTFGVPYNPQS
QGVVEAMNHHLKNQIDRIREQANSVETIVLMAVHCMNFKRRGGIGDMTPAERLINMIT
TEQEIQFQQSKNSKFKNFRVYYREGRDQLWKGPGELLWKGEGAVILKVGTDIKVVPRR
KAKIIKDYGGGKEVDSSSHMEDTGEAREVA"
misc_feature 5084. .5099
/gene="pol"
/note="polypurine tract"
gene 5340. .5984
/gene="vif"
CDS 5340. .5984
/gene="vif"
/codon_start=1
/product="vif protein"
/protein_id="AAK64204.1"
/db_xref="GI:14600352"
/translation="MEEEKRWIAVPTWRIPERLERWHSLIKYLKYKTKDLQKVCYVPH
FKVGWAWWTCSRVIFPLQEGSHLEVQGYWHLTPEKGWLSTYAVRITWYSKNFWTDVTP
NYADILLHSTYFPCFTAGEVRRAIRGEQLLSCCRFPRAHKYQVPSLQYLALKVVSDVR
SQGENPTWKQWRRDNRRGLRMAKQNSRGDKQRGGKPPTKGANFPGLAKVLGILA"
gene 5812. .6150
/gene="vpx"
CDS 5812. .6150
/gene="vpx"
/codon_start=1
/product="vpx protein"
/protein_id="AAK64205.1"
/db_xref="GI:14600353"
/translation="MSDPRERIPPGNSGEETIGEAFEWLNRTVEEINREAVNHLPREL
IFQVWQRSWEYWHDEQGMSPSYVKYRYLCLIQKALFMHCKKGCRCLGEGHGAGGWRPG
PPPPPPPGLA"
gene 6151. .6456
/gene="vpr"

```
     CDS             6151. .6456
                     /gene="vpr"
                     /codon_start=1
                     /product="vpr protein"
                     /protein_id="AAK64206.1"
                     /db_xref="GI:14600354"
                     /translation="MEERPPENEGPQREPWDEWVVEVLEELKEEALKHFDPRLLTALG
                     NHIYNRHGDTLEGAGELIRILQRALFMHFRGGCIHSRIGQPGGGNPLSAIPPSRSML"
     gene            6302. .8902
                     /gene="tat"
     CDS             join(6302. .6597,8803. .8902)
                     /gene="tat"
                     /codon_start=1
                     /product="tat protein"
                     /protein_id="AAK64199.1"
                     /db_xref="GI:14600347"
                     /translation="METPLRQEQENSLESSNERSSCISEADASTPESANLGEEILSQLY
                     RPLEACYNTCYCKKCCYHCQFCFLKKGLGICYEQSRKRRRTPKKAKANTSSASNNRPI
                     SNRTRHCQPEKAKKETVEKAVATAPGLGR"
     gene            6528. .9059
                     /gene="rev"
     CDS             join(6528. .6597,8803. .9059)
                     /gene="rev"
                     /codon_start=1
                     /product="rev protein"
                     /protein_id="AAK64200.1"
                     /db_xref="GI:14600348"
                     /translation="MNNHEREEELRKRLRLIHLLHQTTDPYPTGPGTANQRRQRKRRW
                     RRRWQQLLALADRIYSFPDPPTDTPLDLAIQQLQNLAIESIPDPPTNTPEALCDPTED
                     SRSPQD"
     gene            6604. .9243
                     /gene="env"
     CDS             6604. .9243
                     /gene="env"
                     /note="envelope glycoprotein; includes both SU and TM
                     portions"
                     /codon_start=1
                     /product="env protein"
                     /protein_id="AAK64201.1"
                     /db_xref="GI:14600349"
                     /translation="MGCLGNQLLIAILLLSVYGIYCTLYVTVFYGVPVWRNATIPLFC
                     ATKNRDTWGTTQCLPDNGDYSEMALNVTESFDAWNNTVTEQAIEDVWQLFETSIKPCV
                     KLSPLCITMRCNKSETDRWGLTKSITTTASTTSTTASARVDMVNETSSCIAQDNCAGL
                     EQEQMISCKFNMTGLNRDKKKEYNETWYSADLVCEQGNNTGNESRCYMNHCNTSVIKE
                     SCDKHYWDAIRFRYCAPPGYALLRCNDTNYSGFMPKCSKVVVSSCTRMMETQTSTWFG
                     FNGTRAENRTYIYWHGRDNRTIISLNKYYNLTIKCRRPGNKTVLPVTIMSGLVFHSQR
                     INDRPRQAWCWFGGKWKDAIKEVKQTIVKHPRYTGTNNTDKINLTAPRGGDPEVTFMW
                     TNCRGEFLYCKMNWFLNWVEDRNTANQKPKEQHKRNYVPCHIRQIINTWHKVGKNVYL
                     PPREGDLTCNSTVTSLIANIDWIDGNQTNITMSAEVAELYRLELGDYKLVEITPIGLA
                     PTDVKRYTTGGTSRNKRGVFVLGFLGFLATAGSAMGAASLTLTAQSRTLLAGIVQQQQ
                     QLLDVVKRQQELLRLTVWGTKNLQTRVTAIEKYLKDQAQLNTWGCAFRQVCHTTVPWP
                     NASLIPKWNNETWQEWERKVDFLEENITALLEEAQIQQEKNMYELQKLNSWNVFGNWF
                     DLASWIKYIQYGVYIVVGVILLRIVIYIVQVLAKLRQGYRPVFSSPPSYFQQTHIQQD
                     PALPTREGKERDGGEGGGNSSWPWQIEYIHFLIRQLIRLLTWLFSNCRTLLSRVYQIL
                     QPILQRLSATLQRIREVLRTELTYLQYGWSYFHEAVQAVWRSATETLAGAWGDLWETL
                     RRGGRWILAIPRRIRQGLELTLL"
     gene            9077. .9868
                     /gene="nef"
     CDS             /note="contains premature stop codon"
                     9077. .9355
                     /gene="nef"
                     /codon_start=1
                     /product="truncated nef protein"
                     /protein_id="AAK64207.1"
                     /db_xref="GI:14600355"
                     /translation="MGGAISMRRSRPSGDLRQRLLRARGETYGRLLGEVEDGYSQSPG
     misc_feature    GLDKGLSSLSCEGQKYNQGQYMNTPWRNPAEEREKLAYRKQNMDDIDE"
                     9445. .9460
                     /gene="nef"
     LTR             /note="polypurine tract"
                     9462. .10279
     misc_feature    /note="3' long terminal repeat"
                     9462. .9978
                     /note="U3 region"
     misc_feature    9979. .10155
                     /note="R region"
     polyA_signal    10132. .10137
     misc_feature    10156. .10279
                     /note="U5 region"
BASE COUNT      3468 a    1932 c    2569 g    2310 t
ORIGIN

  Query Match             100.0%;  Score 1913;  DB 14;  Length 10279;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1913;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GGTACCAGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGCAAC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     937 GGTACCAGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGCAAC 996

Qy      61 ACAAAAAAGAAATAGCTGTCTTTTATCCAGGAAGGGGTAATAAGATAGAGTGGGAGATGG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     997 ACAAAAAAGAAATAGCTGTCTTTTATCCAGGAAGGGGTAATAAGATAGAGTGGGAGATGG 1056

Qy     121 GCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCTAC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1057 GCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCTAC 1116

Qy     181 GACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTAG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1117 GACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTAG 1176

Qy     241 ATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTTT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1177 ATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTTT 1236

Qy     301 CGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACTG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1237 CGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACTG 1296

Qy     361 TCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAAC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1297 TCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAAC 1356

Qy     421 AGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAAACAA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1357 AGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAAACAA 1416

Qy     481 GTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGGTG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1417 GTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGGTG 1476

Qy     541 GTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGATAG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1477 GTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGATAG 1536

Qy     601 AGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGTTGCA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1537 AGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGTTGCA 1596

Qy     661 CCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCAGA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1597 CCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCAGA 1656

Qy     721 TTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACCAG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1657 TTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACCAG 1716

Qy     781 CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACTAGTT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1717 CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACTAGTT 1776

Qy     841 CAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGCAACA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1777 CAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGCAACA 1836
```

```
Qy     901 TTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCAACAA 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1837 TTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCAACAA 1896

Qy     961 ACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCT 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1897 ACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCT 1956

Qy    1021 ACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACAC 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1957 ACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACAC 2016

Qy    1081 TGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATC 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2017 TGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATC 2076

Qy    1141 CCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCTA 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2077 CCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCTA 2136

Qy    1201 GATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCAGCAG 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2137 GATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCAGCAG 2196

Qy    1261 CCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGACACT 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2197 CCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGACACT 2256

Qy    1321 CTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATGGACC 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2257 CTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATGGACC 2316

Qy    1381 ATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGGGGAA 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2317 ATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGGGGAA 2376

Qy    1441 AGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGCTCCCC 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2377 AGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGCTCCCC 2436

Qy    1501 CAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAGAGAG 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2437 CAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAGAGAG 2496

Qy    1561 AAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCACC 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2497 AAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCACC 2556

Qy    1621 TCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAGA 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2557 TCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAGA 2616

Qy    1681 AGTATTACTGGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCACA 1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2617 AGTATTACTGGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCACA 2676

Qy    1741 TTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAAA 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2677 TTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAAA 2736

Qy    1801 TGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCCC 1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2737 TGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCCC 2796

Qy    1861 GATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAATT 1913
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2797 GATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAATT 2849


RESULT 7
SIVMM239
LOCUS       SIVMM239               10535 bp ss-RNA     linear   VRL 18-APR-2000
DEFINITION  Simian (macaque) immunodeficiency virus, isolate 239, complete
            proviral genome and flanking sequence.
ACCESSION   M33262
VERSION     M33262.1  GI:334647
KEYWORDS    .
SOURCE      Simian immunodeficiency virus.
  ORGANISM  Simian immunodeficiency virus
            Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
            lentivirus group.
REFERENCE   1  (bases 1 to 10535)
  AUTHORS   Kestler,H., Kodama,T., Ringler,D., Marthas,M., Pedersen,N.C.,
            Lackner,A., Regier,D., Sehgal,P., Daniel,M., King,N. and
            Desrosiers,R.
  JOURNAL   Unpublished
REFERENCE   2  (sites)
  AUTHORS   Kestler,H., Kodama,T., Ringler,D., Marthas,M., Pedersen,N.C.,
            Lackner,A., Regier,D., Sehgal,P., Daniel,M., King,N. and
            Desrosiers,R.
  TITLE     Induction of AIDS in rhesus monkeys by molecularly cloned simian
            immunodeficiency virus
  JOURNAL   Science 248 (4959), 1109-1112 (1990)
  MEDLINE   90260657
   PUBMED   2160735
COMMENT     [2]  sites; for [1].
            Kindly submitted in computer-readable form prior to publication by
            Dr. Desrosiers, New England Regional Primate Center, Harvard.  In
            addition to the 256 bases of flanking cellular DNA shown below,
            approximately 2500 bases of 3' flanking material were determined.
               Author address R.Desrosiers,
               New England Regional Primate Research Center
               Harvard Medical School
               Southborough MA 01772.
FEATURES             Location/Qualifiers
     source          1. .10535
                     /organism="Simian immunodeficiency virus"
                     /isolate="239"
                     /db_xref="taxon:11723"
     LTR             257. .1074
                     /note="5'LTR"
     LTR             257. .773
                     /note="U3"
     repeat_region   774. .950
                     /note="R"
     misc_feature    775
                     /note="mRNA start (putative); putative"
     LTR             951. .1074
                     /note="U5"
     CDS             1309. .2841
                     /note="gag polyprotein"
                     /codon_start=1
                     /protein_id="AAA47632.1"
                     /db_xref="GI:334650"
                     /translation="MGVRNSVLSGKKADELEKIRLRPNGKKKYMLKHVVWAANELDRF
                     GLAESLLENKEGCQKILSVLAPLVPTGSENLKSLYNTVCVIWCIHAEEKVKHTEEAKQ
                     IVQRHLVVETGTTETMPKTSRPTAPSSGRGGNYPVQQIGGNYVHLPLSPRTLNAWVKL
                     IEEKKFGAEVVPGFQALSEGCTPYDINQMLNCVGDHQAAMQIIRDIINEEAADWDLQH
                     PQPAPQQGQLREPSGSDIAGTTSSVDEQIQWMYRQQNPIPVGNIYRRWIQLGLQKCVR
                     MYNPTNILDVKQGPKEPFQSYVDRFYKSLRAEQTDAAVKNWMTQTLLIQNANPDCKLV
                     LKGLGVNPTLEEMLTACQGVGGPGQKARLMAEALKEALAPVPIPFAAAQQRGPRKPIK
                     CWNCGKEGHSARQCRAPRRQGCWKCGKMDHVMAKCPDRQAGFLGLGPWGKKPRNFPMA
                     QVHQGLMPTAPPEDPAVDLLKNYMQLGKQQREKQRESREKPYKEVTEDLLHLNSLFGG
                     DQ"
     CDS             <2484. .5666
                     /note="pol polyprotein"
                     /codon_start=1
                     /protein_id="AAA47633.2"
                     /db_xref="GI:7583329"
                     /translation="VLELWERGTLCKAMQSPKKTGMLEMWKNGPCYGQMPRQTGGFFR
                     PWSMGKEAPQFPHGSSASGADANCSPRGPSCGSAKELHAVGQAAERKAERKQREALQG
                     GDRGFAAPQFSLWRRPVVTAHIEGQPVEVLLDTGADDSIVTGIELGPHYTPKIVGGIG
                     GFINTKEYKNVEIEVLGKRIKGTIMTGDTPINIFGRNLLTALGMSLNFPIAKVEPVKV
                     ALKPGKDGPKLKQWPLSKEKIVALREICEKMEKDGQLEEAPPTNPYNTPTFAIKKKDK
                     NKWRMLIDFRELNRVTQDFTEVQLGIPHPAGLAKRKRITVLDIGDAYFSIPLDEEFRQ
                     YTAFTLPSVNNAEPGKRYIYKVLPQGWKGSPAIFQYTMRHVLEPFRKANPDVTLVQYM
```

```
                     DDILIASDRTDLEHDRVVLQSKELLNSIGFSTPEEKFQKDPPFQWMGYELWPTKWKLQ
                     KIELPQRETWTVNDIQKLVGVLNWAAQIYPGIKTKHLCRLIRGKMTLTEEVQWTEMAE
                     AEYEENKIILSQEQEGCYYQEGKPLEATVIKSQDNQWSYKIHQEDKILKVGKFAKIKN
                     THTNGVRLLAHVIQKIGKEAIVIWGQVPKFHLPVEKDVWEQWWTDYWQVTWIPEWDFI
                     STPPLVRLVFNLVKDPIEGEETYYTDGSCNKQSKEGKAGYITDRGKDKVKVLEQTTNQ
                     QAELEAFLMALTDSGPKANIIVDSQYVMGIITGCPTESESRLVNQIIEEMIKKSEIYV
                     AWVPAHKGIGGNQEIDHLVSQGIRQVLFLEKIEPAQEEHDKYHSNVKELVFKFGLPRI
                     VARQIVDTCDKCHQKGEAIHGQANSDLGTWQMDCTHLEGKIIIVAVHVASGFIEAEVI
                     PQETGRQTALFLLKLAGRWPITHLHTDNGANFASQEVKMVAWWAGIEHTFGVPYNPQS
                     QGVVEAMNHHLKNQIDRIREQANSVETIVLMAVHCMNFKRRGGIGDMTPAERLINMIT
                     TEQEIQFQQSKNSKFKNFRVYYREGRDQLWKGPGELLWKGEGAVILKVGTDIKVVPRR
                     KAKIIKDYGGGKEVDSSSHMEDTGEAREVA"
misc_feature         5340. .5355
                     /note="polypurine tract"
CDS                  5596. .6240
                     /note="vif protein"
                     /codon_start=1
                     /protein_id="AAA47634.1"
                     /db_xref="GI:334652"
                     /translation="MEEEKRWIAVPTWRIPERLERWHSLIKYLKYKTKDLQKVCYVPH
                     FKVGWAWWTCSRVIFPLQEGSHLEVQGYWHLTPEKGWLSTYAVRITWYSKNFWTDVTP
                     NYADILLHSTYFPCFTAGEVRRAIRGEQLLSCCRFPRAHKYQVPSLQYLALKVVSDVR
                     SQGENPTWKQWRRDNRRGLRMAKQNSRGDKQRGGKPPTKGANFPGLAKVLGILA"
CDS                  6068. .6406
                     /note="vpX protein"
                     /codon_start=1
                     /protein_id="AAA47635.1"
                     /db_xref="GI:334653"
                     /translation="MSDPRERIPPGNSGEETIGEAFEWLNRTVEEINREAVNHLPREL
                     IFQVWQRSWEYWHDEQGMSPSYVKYRYLCLIQKALFMHCKKGCRCLGEGHGAGGWRPG
                     PPPPPPPGLA"
CDS                  6407. .6712
                     /note="vpR protein"
                     /codon_start=1
                     /protein_id="AAA47636.1"
                     /db_xref="GI:334654"
                     /translation="MEERPPENEGPQREPWDEWVVEVLEELKEEALKHFDPRLLTALG
                     NHIYNRHGDTLEGAGELIRILQRALFMHFRGGCIHSRIGQPGGGNPLSAIPPSRSML"
CDS                  join(6558. .6853,9062. .9158)
                     /note="tat protein"
                     /codon_start=1
                     /protein_id="AAA47630.1"
                     /db_xref="GI:334648"
                     /translation="METPLREQENSLESSNERSSCISEADASTPESANLGEEILSQLY
                     RPLEACYNTCYCKKCCYHCQFCFLKKGLGICYEQSRKRRRTPKKAKANTSSASNKPIS
                     NRTRHCQPEKAKKETVEKAVATAPGLGR"
exon                 <6558. .6853
                     /note="tat protein,  (first expressed exon)"
                     /number=2
CDS                  join(6784. .6853,9062. .9315)
                     /note="rev protein"
                     /codon_start=1
                     /protein_id="AAA47631.1"
                     /db_xref="GI:334649"
                     /translation="MSNHEREEELRKRLRLIHLLHQTNPYPTGPGTANQRRQRKRRWR
                     RRWQQLLALADRIYSFPDPPTDTPLDLAIQQLQNLAIESIPDPPTNTPEALCDPTEDS
                     RSPQD"
exon                 <6784. .6853
                     /note="rev protein,  (first expressed exon)"
                     /number=2
CDS                  6860. .9499
                     /note="env polyprotein"
                     /codon_start=1
                     /protein_id="AAA47637.1"
                     /db_xref="GI:334655"
                     /translation="MGCLGNQLLIAILLLSVYGIYCTLYVTVFYGVPAWRNATIPLFC
                     ATKNRDTWGTTQCLPDNGDYSEVALNVTESFDAWNNTVTEQAIEDVWQLFETSIKPCV
                     KLSPLCITMRCNKSETDRWGLTKSITTTASTTSTTASAKVDMVNETSSCIAQDNCTGL
                     EQEQMISCKFNMTGLKRDKKKEYNETWYSADLVCEQGNNTGNESRCYMNHCNTSVIQE
                     SCDKHYWDAIRFRYCAPPGYALLRCNDTNYSGFMPKCSKVVVSSCTRMMETQTSTWFG
                     FNGTRAENRTYIYWHGRDNRTIISLNKYYNLTMKCRRPGNKTVLPVTIMSGLVFHSQP
                     INDRPKQAWCWFGGKWKDAIKEVKQTIVKHPRYTGTNNTDKINLTAPGGGDPEVTFMW
                     TNCRGEFLYCKMNWFLNWVEDRNTANQKPKEQHKRNYVPCHIRQIINTWHKVGKNVYL
                     PPREGDLTCNSTVTSLIANIDWIDGNQTNITMSAEVAELYRLELGDYKLVEITPIGLA
                     PTDVKRYTTGGTSRNKRGVFVLGFLGFLATAGSAMGAASLTLTAQSRTLLAGIVQQQQ
                     QLLDVVKRQQELLRLTVWGTKNLQTRVTAIEKYLKDQAQLNAWGCAFRQVCHTTVPWP
                     NASLTPKWNNETWQEWERKVDFLEENITALLEEAQIQQEKNMYELQKLNSWDVFGNWF
                     DLASWIKYIQYGVYIVVGVILLRIVIYIVQMLAKLRQGYRPVFSSPPSYFQQTHIQQD
                     PALPTREGKERDGGEGGGNSSWPWQIEYIHFLIRQLIRLLTWLFSNCRTLLSRVYQIL
                     QPILQRLSATLQRIREVLRTELTYLQYGWSYFHEAVQAVWRSATETLAGAWGDLWETL
                     RRGGRWILAIPRRIRQGLELTLL"
exon                 9062. .>9315
                     /note="rev protein"
                     /number=3
exon                 9062. .>9158
                     /note="tat protein"
                     /number=3
CDS                  9333. .10124
                     /note="nef protein (in-frame stop at 9612)"
                     /pseudo
                     /codon_start=1
CDS                  9333. .9611
                     /note="nef protein (premature termination)"
                     /codon_start=1
                     /protein_id="AAA47638.1"
                     /db_xref="GI:334656"
                     /translation="MGGAISMRRSRPSGDLRQRLLRARGETYGRLLGEVEDGYSQSPG
                     GLDKGLSSLSCEGQKYNQGQYMNTPWRNPAEEREKLAYRKQNMDDIDE"
misc_feature         9701. .9716
                     /note="polypurine tract"
LTR                  9718. .10535
                     /note="3' LTR"
LTR                  9718. .10234
                     /note="U3"
repeat_region        10235. .10411
                     /note="R"
polyA_signal         10388. .10393
LTR                  10412. .10535
                     /note="U5"
BASE COUNT     3540 a   1988 c   2607 g   2400 t
ORIGIN

  Query Match           100.0%;  Score 1913;  DB 14;  Length 10535;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1913;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GGTACCAGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGCAAC 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1193 GGTACCAGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGCAAC 1252

Qy      61 ACAAAAAAGAAATAGCTGTCTTTTATCCAGGAAGGGGTAATAAGATAGAGTGGGAGATGG 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1253 ACAAAAAAGAAATAGCTGTCTTTTATCCAGGAAGGGGTAATAAGATAGAGTGGGAGATGG 1312

Qy     121 GCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCTAC 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1313 GCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCTAC 1372

Qy     181 GACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTAG 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1373 GACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTAG 1432

Qy     241 ATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTTT 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1433 ATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTTT 1492

Qy     301 CGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACTG 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1493 CGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACTG 1552

Qy     361 TCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAAC 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1553 TCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAAC 1612

Qy     421 AGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAAACAA 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1613 AGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAAACAA 1672

Qy    481 GTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGGTG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1673 GTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGGTG 1732

Qy    541 GTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGATAG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1733 GTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGATAG 1792

Qy    601 AGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGTTGCA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1793 AGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGTTGCA 1852

Qy    661 CCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCAGA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1853 CCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCAGA 1912

Qy    721 TTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACCAG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1913 TTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACCAG 1972

Qy    781 CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACTAGTT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1973 CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACTAGTT 2032

Qy    841 CAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGCAACA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2033 CAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGCAACA 2092

Qy    901 TTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCAACAA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2093 TTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCAACAA 2152

Qy    961 ACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2153 ACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCT 2212

Qy   1021 ACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACAC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2213 ACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACAC 2272

Qy   1081 TGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATC 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2273 TGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATC 2332

Qy   1141 CCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCTA 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2333 CCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCTA 2392

Qy   1201 GATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCAGCAG 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2393 GATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCAGCAG 2452

Qy   1261 CCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGACACT 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2453 CCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGACACT 2512

Qy   1321 CTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATGGACC 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2513 CTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATGGACC 2572

Qy   1381 ATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGGGGAA 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2573 ATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGGGGAA 2632

Qy   1441 AGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGCTCCCC 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2633 AGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGCTCCCC 2692

Qy   1501 CAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAGAGAG 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2693 CAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAGAGAG 2752

Qy   1561 AAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCACC 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2753 AAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCACC 2812

Qy   1621 TCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAGA 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2813 TCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAGA 2872

Qy   1681 AGTATTACTGGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCACA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2873 AGTATTACTGGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCACA 2932

Qy   1741 TTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAAA 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2933 TTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAAA 2992

Qy   1801 TGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCCC 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2993 TGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCCC 3052

Qy   1861 GATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAATT 1913
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3053 GATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAATT 3105
```

```
RESULT 8
AF217181
LOCUS       AF217181               9860 bp    RNA     linear   VRL 01-JAN-2002
DEFINITION  Simian-Human immunodeficiency virus SHIV-C2/1, complete genome.
ACCESSION   AF217181
VERSION     AF217181.1  GI:18026826
KEYWORDS    .
SOURCE      Simian-Human immunodeficiency virus.
  ORGANISM  Simian-Human immunodeficiency virus
            Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
            lentivirus group.
REFERENCE   1  (bases 1 to 9860)
  AUTHORS   Sakai,K., Shinohara,K., Takahashi,E., Ami,Y., Sasaki,Y., Suzaki,Y.,
            Nakasone,T. and Honda,M.
  TITLE     Comparative genetic analysis of simian/human immunodeficiency
            viruses (SHIVs) with pathogenic property
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 9860)
  AUTHORS   Sakai,K., Shinohara,K., Takahashi,E. and Honda,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1999) AIDS Research Center, National Institute of
            Infectious Diseases, Gakuen 4-7-1, Musashimurayama, Tokyo 208-0011,
            Japan
FEATURES             Location/Qualifiers
     source          1. .9860
                     /organism="Simian-Human immunodeficiency virus"
                     /virion
                     /strain="SHIV-C2/1"
                     /specific_host="cynomolgus monkey"
                     /db_xref="taxon:57667"
                     /note="obtained from monkey plasma"
     misc_feature    1. .5938
                     /note="derived from SIVmac239"
     LTR             1. .301
     repeat_region   1. .177
                     /note="R region"
     misc_feature    178. .301
                     /note="U5 region"
     gene            536. .2068
                     /gene="gag"
     CDS             536. .2068
                     /gene="gag"
                     /codon_start=1
                     /product="gag protein"
                     /protein_id="AAL55638.1"
                     /db_xref="GI:18026827"
                     /translation="MGVRNSVLSGKKADELEKIRLRPNGKKKYMLKHVVWAANELDRF
```

```
                GLAESLLENKEGCQKILSVLAPLVPTGSENLKSLYNTVCVIWCIHAEEKVKHTEEAKQ
                IVQRHLVVETGTTETMPKTSRPTAPSSGRGGNYPVQQIGGNYVHLPLSPRTLNAWVKL
                IEEKKFGAEVVPGFQALSEGCTPYDINQMLNCVGDHQAAMQIIRDIINEEAADWDLQH
                PQPAPQQGQLREPSGSDIAGTTSSVDEQIQWMYRQQNPIPVGNIYRRWIQLGLQKCVR
                MYNPTNILDVKQGPKEPFQSYVDRFYKSLRAEQTDAAVKNWMTQTLLIQNANPDCKLV
                LKGLGVNPTLEEMLTACQGVGGPGQKARLMAEALKEALAPVPIPFAAAQQRGPRKPIK
                CWNCGKEGHSARQCRAPRRQGCWKCGKMDHVMAKCPDRQAGFLGLGPWGKKPRNFPMA
                QVHQGLIPTAPPEDPAVDLLKNYMQLGKQQREKQRESREKPYKEVTEDLLHLNSLFGG
                DQ"
gene            <1711. .4893
                /gene="pol"
CDS             <1711. .4893
                /gene="pol"
                /codon_start=1
                /product="pol protein"
                /protein_id="AAL55639.1"
                /db_xref="GI:18026828"
                /translation="VLELWERGTLCKAMQSPKKTGMLEMWKNGPCYGQMPRQTGGFFR
                PWSMGKEAPQFPHGSSASGADTNCSPRGPSCGSAKELHAVGQAAERKAERKQREALQG
                GDRGFAAPQFSLWRRPVVTAHIEGQPVEVLLDTGADDSIVTGIELGPHYTPKIVGGIG
                GFINTKEYKNVEIEVLGKRIKGTIMTGDTPINIFGRNLLTALGMSLNFPIAKVEPVKV
                ALKPGKDGPKLKQWPLSKEKIVALREICEKMEKDGQLEEAPPTNPYNTPTFAIKKKDK
                NKWRMLIDFRELNRVTQDFTEVQLGIPHPAGLAKRKRITVLDIGDAYFSIPLDEEFRQ
                YTAFTLPSVNNAEPGKRYIYKVLPQGWKGSPAIFQYTMRHVLEPFRKANPDVTLVQYM
                DDILIASDRTDLEHDRVVLQLKELLNSIGFSTPEEKFQKDPPFQWMGYELWPTKWKLQ
                KIELPQRETWTVNDIQKLVGVLNWAAQIYPGIKTKHLCRLIRGKMTLTEEVQWTEMAE
                AEYEENKIILSQEQEGCYYQEGKPLEATVIKSQDNQWSYKIHQEDKILKVGKFAKIKN
                THTNGVRLLAHVIQKIGKEAIVIWGQVPKFHLPVEKDVWEQWWTDYWQVTWIPEWDFI
                STPPLVRLVFNLVKDPIEGEETYYTDGSCNKQSKEGKAGYITDRGKDKVKVLEQTTNQ
                QAELEAFLMALTDSGPKANIIVDSQYVMGIITGCPTESESRLVNQIIEEMIKKSEIYV
                AWVPAHKGIGGNQEIDHLVSQGIRQVLFLERIEPAQEEHDKYHSNVKELVFKFGLPRI
                VARQIVDTCDKCHQKGEAIHGQANSDLGTWQMDCTHLEGKIIIVAVHVASGFIEAEVI
                PQETGRQTALFLLKLAGRWPITHLHTDNGANFASQEVKMVAWWAGIEHTFGVPYNPQS
                QGVVEAMNHHLKNQIDRIREQANSVETIVLMAVHCMNFKRRGGIGDMTPAERLINMIT
                TEQEIQFQQSKNSKFKNFRVYYREGRDQLWKGPGELLWKGEGAVILKVGTDIKVVPRR
                KAKIIKDYGGGKEVDSSSHMEDTGEAREVA"
gene            4823. .5467
                /gene="vif"
CDS             4823. .5467
                /gene="vif"
                /codon_start=1
                /product="vif protein"
                /protein_id="AAL55640.1"
                /db_xref="GI:18026829"
                /translation="MEEEKRWIAVPTWRIPERLERWHSLIKYLKYKTKDLQKVCYVPH
                FKVGWAWWTCSRVIFPLQEGSHLEVQGYWHLTPEKGWLSTYAVRITWYSKNFWTDVTP
                NYADILLHSTYFPCFTAGEVRRAIRGEQLLSCCRFPRAHKYQVPSLQYLALKVVSDVR
                SQGENPTWKQWRRDNRRGLRMAKQNSRGDKQRGGKPPTKGANFPGLAKVLGILA"
gene            5295. .5633
                /gene="vpx"
CDS             5295. .5633
                /gene="vpx"
                /codon_start=1
                /product="vpx protein"
                /protein_id="AAL55641.1"
                /db_xref="GI:18026830"
                /translation="MSDPRERIPPGNSGEETIGEAFEWLNRTVEEINREAVNHLPREL
                IFQVWQRSWEYWHDEQGMSPSYVKYRYLCLIQKALFMHCKKGCRCLGEGHGAGGWRPG
                PPPPPPPGLA"
gene            5634. .5939
                /gene="vpr"
CDS             5634. .5939
                /gene="vpr"
                /codon_start=1
                /product="vpr protein"
                /protein_id="AAL55642.1"
                /db_xref="GI:18026831"
                /translation="MEERPPENEGPQREPWDEWVVEVLEELKEEALKHFDPRLLTALG
                NHIYNRHGDTLEGAGELIRILQRALFMHFRGGCIHSRIGQPGGGNPLSAIPPSRSML"
misc_feature    5939. .8775
                /note="derived from HIV-1"
gene            5948. .8580
                /gene="tat"
CDS             join(5948. .6162,8487. .8580)
                /gene="tat"
                /codon_start=1
                /product="tat protein"
                /protein_id="AAL55643.1"
                /db_xref="GI:18026832"
                /translation="MEPVDPRLEPWKHPGSKPKTACTNCYCKKCCFHCQVCFTTKALG
                ISYGRKKRRQRRRAHQNSQTHQASLSKQPSSQPRGDPTGPKEQKKKVERETETDPVHQ
                "
gene            6087. .8761
                /gene="rev"
CDS             join(6087. .6162,8487. .8761)
                /gene="rev"
                /codon_start=1
                /product="rev protein"
                /protein_id="AAL55644.1"
                /db_xref="GI:18026833"
                /translation="MAGRSGDSDEELIRTVRLIKLLYQSNPPPSLEGTRQARRNRRRR
                WRERQRQIRSISERILGTYLGRSAEPVPLQLPPLERLTLDCNEDCGTSGTQGVGSPQI
                LVESPTVLESGTKE"
gene            6179. .6424
                /gene="vpu"
CDS             6179. .6424
                /gene="vpu"
                /codon_start=1
                /product="vpu protein"
                /protein_id="AAL55645.1"
                /db_xref="GI:18026834"
                /translation="MQPIQIAIVALVVAIIIAIVVWSIVIIEYRKILRQRKIDRLIDR
                LIERAEDSGNESEGEISALVEMGVEMGHHAPWDVDDL"
gene            6339. .8948
                /gene="env"
CDS             6339. .8948
                /gene="env"
                /codon_start=1
                /product="env polyprotein"
                /protein_id="AAL55646.1"
                /db_xref="GI:18026835"
                /translation="MRVKEKYQHLWRWGWRWGTMLLGMLMICSATEKLWVTVYYGVPV
                WREATTTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEVVLGNVTENFNMWKNNMVD
                QMHEDIISLWDESLKPCVKLTPLCVTLNCTNLNITKNTTNLTSSSWGMMEEGEIKNCS
                FYITTSIRNKVKKEYALFNRLDVVPVKNTSNTKYRLISCNTSVITQACPKVSFQPIPI
                HYCVPAGFAILKCNNKTFNGSGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAEEDIVIR
                SEDFTDNVKTIIVQLNESVVINCTRPNNNTRERLSIGPGRAFYARRNIIGDIRQAHCN
                ISRAKWNNTLQQIVIKLREKFRNKTIAFNQSSGGDPEIVMHSFNCGGEFFYCNTAQLF
                NSTWNVAGGTNGTEGNDIITLQCRIKQIINMWQKVGKAMYAPPITGQIRCSSNITGLL
                LTRDGGNSTETETEIFRPGGGDMRDNWRSELYKYKVVRIEPIGVAPTRAKRRTVQREK
                RAVGIGAVFLGFLGAAGSTMGAASVTLTVQARLLLSGIVQQQNNLLRAIEAQQNMLRL
                TVWGIKQLQARVLALERYLRDQQLMGIWGCSGKLICTTSVPWNVSWSNKSVDDIWNNM
                TWMEWEREIDNYTDYIYDLLEKSQTQQEKNEKELLELDKWASLWNWFDITNWLWYIRL
                FIMIVGGLIGLRIVFAVLSIVNRVRQGYSPLSFQTLLPASRGPDRPEGTEEEGGERDR
                DRSGPSVNGSLALIWADLRSLCLFSYHRLRDLLLIVTRTVELLGRRGWEALKYWWNLL
                QYWSQELKNSAVSLLQYGWSYFHEAVQAVWRSATETLAGAWGDVWETLRRGGRWILAI
                PRRIRQGLELTLL"
misc_feature    8776. .9860
                /note="derived from SIVmac239"
gene            8782. .9573
                /gene="nef"
CDS             8782. .9573
                /gene="nef"
                /codon_start=1
                /product="nef protein"
                /protein_id="AAL55647.1"
                /db_xref="GI:18026836"
                /translation="MGGAISMRRSRPSGDLRQRLLRARGETYGRLLGEVEDGYSQSPG
                GLDKGLSSLSCEGQKYNQGQYMNTPWRNPAEEREKLAYRKQNMDDIDEEDDDLVGVSV
                RPKVPLKTMSYKLAIDMSHFIKEKGGLEGIYYSARRHRILDIYLEKEEGIIPDWQDYT
                SGPGIRYPKTFGWLWKLVPVNVSDEAQEDEEHYLMHPAQTSQWDDPWGEVLAWKFDPT
                LAYTYEAYVRYPEEFGSKSGLSEEEVRRRLTARGLLNMADKKETR"
LTR             9167. .9860
misc_feature    9167. .9683
                /note="U3 region"
repeat_region   9684. .9860
```

```
                     /note="R region"
BASE COUNT     3415 a   1789 c   2466 g   2190 t
ORIGIN

  Query Match             99.8%;  Score 1909.8;  DB 14;  Length 9860;
  Best Local Similarity   99.9%;  Pred. No: 0;
  Matches 1911;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GGTACCAGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGCAAC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     420 GGTACCAGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGCAAC 479

Qy      61 ACAAAAAAGAAATAGCTGTCTTTTATCCAGGAAGGGGTAATAAGATAGAGTGGGAGATGG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     480 ACAAAAAAGAAATAGCTGTCTTTTATCCAGGAAGGGGTAATAAGATAGAGTGGGAGATGG 539

Qy     121 GCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCTAC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     540 GCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCTAC 599

Qy     181 GACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTAG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     600 GACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTAG 659

Qy     241 ATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTTT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     660 ATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTTT 719

Qy     301 CGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACTG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     720 CGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACTG 779

Qy     361 TCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAAC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     780 TCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAAC 839

Qy     421 AGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAAACAA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     840 AGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAAACAA 899

Qy     481 GTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGGTG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     900 GTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGGTG 959

Qy     541 GTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGATAG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     960 GTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGATAG 1019

Qy     601 AGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGTTGCA 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1020 AGGAAAAGAAATTTGGAGCAGAAGTAGTACCAGGATTTCAGGCACTGTCAGAAGGTTGCA 1079

Qy     661 CCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCAGA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1080 CCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCAGA 1139

Qy     721 TTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACCAG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1140 TTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACCAG 1199

Qy     781 CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACTAGTT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1200 CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACTAGTT 1259

Qy     841 CAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGCAACA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1260 CAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGCAACA 1319

Qy     901 TTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCAACAA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1320 TTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCAACAA 1379

Qy     961 ACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1380 ACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCT 1439

Qy    1021 ACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACAC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1440 ACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACAC 1499

Qy    1081 TGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATC 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1500 TGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATC 1559

Qy    1141 CCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCTA 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1560 CCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCTA 1619

Qy    1201 GATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCAGCAG 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1620 GATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCAGCAG 1679

Qy    1261 CCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGACACT 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1680 CCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGACACT 1739

Qy    1321 CTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATGGACC 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1740 CTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATGGACC 1799

Qy    1381 ATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGGGGAA 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1800 ATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGGGGAA 1859

Qy    1441 AGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGCTCCCC 1500
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    1860 AGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATACCAACTGCTCCCC 1919

Qy    1501 CAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAGAGAG 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1920 CAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAGAGAG 1979

Qy    1561 AAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCACC 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1980 AAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCACC 2039

Qy    1621 TCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAGA 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2040 TCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAGA 2099

Qy    1681 AGTATTACTGGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCACA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2100 AGTATTACTGGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCACA 2159

Qy    1741 TTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAAA 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2160 TTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAAA 2219

Qy    1801 TGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCCC 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2220 TGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCCC 2279

Qy    1861 GATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAATT 1913
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2280 GATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAATT 2332


RESULT 9
SIV1A11AA
LOCUS       SIV1A11AA              10274 bp ss-RNA     linear   VRL 05-MAR-2001
DEFINITION  Simian immunodeficiency virus clone 1A11, complete genome.
ACCESSION   M76764
VERSION     M76764.1  GI:334170
KEYWORDS    .
SOURCE      Simian immunodeficiency virus.
  ORGANISM  Simian immunodeficiency virus
```

```
            Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
            lentivirus group.
REFERENCE   1  (bases 1 to 10274)
  AUTHORS   Marthas,M.L., Banapour,B., Sujipto,S., Siegel,M.E., Marx,P.A.,
            Gardner,M.B. and Luciw,P.A.
  TITLE     Rhesus macaques inoculated with molecularly cloned simian
            immunodeficiency virus
  JOURNAL   J. Med. Primatol. 18, 311-319 (1991)
  MEDLINE   89342414
REFERENCE   2  (bases 1 to 10274)
  AUTHORS   Unger,R.E., Stout,M.W. and Luciw,P.A.
  TITLE     Simian immunodeficiency virus (SIVmac) exhibits complex splicing
            for tat, rev, and env mRNA
  JOURNAL   Virology 182 (1), 177-185 (1991)
  MEDLINE   91220650
   PUBMED   2024463
REFERENCE   3  (bases 1 to 10274)
  AUTHORS   Planelles,V., Haigwood,N.L., Marthas,M.L., Mann,K.A., Scandella,C.,
            Lidster,W.D., Shuster,J.R., Van Kuyk,R., Marx,P.A., Gardner,M.B.
            and Luciw,P.A.
  TITLE     Functional and immunological characterization of SIV envelope
            glycoprotein produced in genetically engineered mammalian cells
  JOURNAL   AIDS Res. Hum. Retroviruses 7 (11), 889-898 (1991)
  MEDLINE   92102765
   PUBMED   1760229
REFERENCE   4  (bases 1 to 10274)
  AUTHORS   Luciw,P.A., Shaw,K.E., Unger,R.E., Planelles,V., Stout,M.W.,
            Lackner,J.E., Pratt-Lowe,E., Leung,N.J., Banapour,B. and
            Marthas,M.L.
  TITLE     Genetic and biological comparisons of pathogenic and nonpathogenic
            molecular clones of simian immunodeficiency virus (SIVmac)
  JOURNAL   AIDS Res. Hum. Retroviruses 8 (3), 395-402 (1992)
  MEDLINE   92239216
   PUBMED   1571198
FEATURES             Location/Qualifiers
     source          1. .10274
                     /organism="Simian immunodeficiency virus"
                     /proviral
                     /db_xref="taxon:11723"
     LTR             1. .817
     TATA_signal     487. .492
                     /note="'R = 518'"
     primer_bind     821. .848
     CDS             1052. .2584
                     /codon_start=1
                     /product="gag polyprotein"
                     /protein_id="AAB54042.1"
                     /db_xref="GI:334173"
                     /translation="MGARNSVLSGKKADELEKIRLRPNGKKKYMLKHVVWAANELDRF
                     GLAESLLENKEGCQKILSVLAPLVPTGSENLKSLYNTVCVIWCIHAEEKVKHTEEAKQ
                     IVQRHLVVETGTAEIMPKTSRPTAPSSGRGGNYPVQQIGGNYVHLPLSPRTLNAWVKL
                     IEEKKFGAEVVPGFQALSEGCTPYDINQMLNCVGDHQAAMQIIRDIINEEAADWDLQH
                     PQPAPQQGQLREPSGSDIAGTTSSVDEQIQWMYRQQNPIPVGNIYRRWIQLGLQKCVR
                     MYNPTNILDVKQGPKEPFQSYVDRFYKSLRAEQTDAAVKNWMTQTLLIQNANPDCKLV
                     LKGLGVNPTLEEMLTACQGVGGPGQKARLMAEALKEALAPVPIPFAAAQQRGPRKPIK
                     CWNCGKEGHSARQCRAPRRQGCWKCGKMDHVMAKCPDRQAGFLGLGPWGKKPRNFPMA
                     QVHQGLTPTAPPEDPAVDLLKNYMQLGKQQREKQRESREKPYKEMTEDLLHLNSLFGG
                     DQ"
     CDS             <2227. .5409
                     /codon_start=1
                     /product="pol polyprotein"
                     /protein_id="AAB54043.1"
                     /db_xref="GI:2065473"
                     /translation="VLELWERGTLCKAMQSPKKTGMLEMWKNGPCYGQMPRQTGGFFR
                     PWSMGKEAPQFPHGSSASGADANCSPRGPSCGSAKELHAVGQAAERKTERKQREALQG
                     DDRGFAAPQFSLWRRPVVTAHIEGQPVEVLLDTGADDSIVTGIELGPHYTPKIVGGIG
                     GFINTKEYKNVEIEVLGKRIRGTIMTGDTPINIFGRNLLTALGMSLNLPIAKVEPVKV
                     ALKPGKVGPKLKQWPLSKEKIVALREICEKMEKDGQLEEAPPTNPYNTPTFAIKKKDK
                     NKWRMLIDFRELNRVTQDFTEVQLGIPHPAGLAKRKRITVLDIGDAYFSIPLDEEFRQ
                     YTAFTLPSVNNAEPGKRYIYKVLPQGWKGSPAIFQYTMRHVLEPFRKANPDVTLVQYM
                     DDILIASDRTNLEHDRVVLQLKELLNSIGFSTPEEKFQKDPPFQWMGYELWPTKWKLQ
                     KIELPQRETWTVNDIQKLVGVLNWAAQIYPGIKTKHLCRLIRGKMTLTEEVQWTEMAE
                     AEYEENKIILSQEQEGCYYQEGKPLEATVIKNQDNQWSYKIHQEDKILKVGKFAKIKN
                     THTNGVRLLAHVIQKIGKEAIVIWGQVPKFHLPVEKDVWEQWWTDYWQVTWIPEWNFI
                     STPPLVRLVFNLVKDPIEGEETYYTDGSCNKQSKEGKAGYITDRGKDKVKVLEQTTNQ
                     QAELEAFLMALTDSGPKANIIVDSQYVMGIITGCPTESESRLVNQIIKEMIKKSEIYV
                     AWVPAHKGIGGNQEVDHLVSQGIRQVLFLEKIEPAQEEHDKYHSNVKELVFKFGLPRI
                     VARQIVDTCDKCHQKGEAIHGQVNSDLGTWQMDCTHLEGKIIIVAVHVASGFIEAEVI
                     PQETGRQTALFLLKLAGRWPITHLHTDNGANFASQEVKMVAWWAGIEHTFGVPYNPQS
                     QGVVEAMNHHLKNQIDRIREQANSVETIVLMAVHSMNFKRRGGTGDMTPAERLINMIT
                     TEQEIQFQQSKNSKFKNFRVYYREGRDQLWKGPGELLWKGEGAVILKVGTDIKVVPRR
                     KAKIIKDYGGGKEMDSSSHMEDTGEAREVA"
     gene            5339. .5983
                     /gene="vif"
     CDS             5339. .5983
                     /gene="vif"
                     /codon_start=1
                     /product="vif protein"
                     /protein_id="AAB54044.1"
                     /db_xref="GI:334175"
                     /translation="MEEEKRWIAVPTWRIPERLERWHSLIKYLKYKTKDLQKVCYVPH
                     FKVGWAWWTCSRVIFPLQEGSHLEVQGYWHLTPEKGWLSTYAVRVTWYSKNFWTDVTP
                     NYADILLHSTYFPCFTAGEVRRAIRGEQLLSCCRFPRAHKYQVPSLQYLALKVVSDVR
                     SQGENPTWKQWRRDNRRGLRMAKQNSRGDKQRGGKPPTKGANFPGLAKVLGILA"
     gene            5811. .6149
                     /gene="vpr"
     CDS             5811. .6149
                     /gene="vpr"
                     /codon_start=1
                     /product="vpr protein"
                     /protein_id="AAB54045.1"
                     /db_xref="GI:334176"
                     /translation="MSDPRERIPPGNSGEETIGEAFEWLNRTVEEINREAVNHLPREL
                     IFQVWQRSWEYWHDEQGMSPSYVKYRYLCLIQKALFMHCKKGCRCLGEGHGAGGWRPG
                     PPPPPPPGLA"
     CDS             6150. .6455
                     /note="'premature stop @ 6185'"
                     /pseudo
     CDS             /codon_start=1
                     join(6301. .6596,8799. .8898)
                     /codon_start=1
                     /product="tat"
                     /protein_id="AAB54040.1"
                     /db_xref="GI:334171"
                     /translation="METPLREQENSLESSSGHSSCTSEAAASTPELANLGEEILSQLY
                     RPLEACYNTCYCKKCCYHCQFCFLKKGLGICYEQSRKRRRTPKKAKANTSSASNNRPI
                     SNRTRHCQPEKAKKETVEKAVATAPGLGR"
     CDS             join(6527. .6596,8799. .9055)
                     /codon_start=1
                     /product="rev"
                     /protein_id="AAB54041.1"
                     /db_xref="GI:334172"
                     /translation="MSSHEREEELRKRLRLIHLLHQTIDPYPTGPGTANQRRQRRRRW
                     RRRWQQLLALADRIYSFPDPPTDTPLDLAIQQLQNLAIESIPDPPTNTPEALCDPTKD
                     SRSPQD"
     CDS             6603. .9239
                     /note="env; 'Accession Number M74144(1 to 1572 = 6603 to
                     8174; SU)'; 'for splice patterns, see Unger,R.E. et al.';
                     'premature stop 1 @ 8801 (SIVmm142)'; 'premature stop 2 @
                     9158'"
                     /pseudo
     CDS             /codon_start=1
                     9073. .9864
                     /codon_start=1
                     /product="nef"
                     /protein_id="AAB54046.1"
                     /db_xref="GI:334177"
                     /translation="MGGTISMRRSRSTGDLRQRLLRARGETYERLLGEVEDGSSQSLG
                     ELDKGLSSLSCEGQKYNQEQYMNTPWRNPAEEREKLAYRKQNMDDIDEEDDDLVGVSV
                     RPKVPLRTMSYKLAIDMSHFIKEKGGLEGIYYSARRHRILDIYLEKEEGIIPDWQDYT
                     SGPGIRYPKTFGWLWKLVPVNVSDEAQEDEEHYLMHPAQTSQWDDPWGEVPAWKFDPT
                     LAYTYEAYVRYPEEFGSKSGLSEEEVRRRLTARGLLNMADKKETR"
     LTR             9458. .10274
     polyA_signal    10127. .10132
                     /note="'670. .675 for the 5'LTR'"
BASE COUNT     3475 a   1931 c   2559 g   2309 t
```

ORIGIN

Query Match 98.5%; Score 1884.8; DB 14; Length 10274;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 1895; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

```
Qy       1 GGTACCAGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGCAAC 60
Db     936 GGTACCAGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGCAAC 995

Qy      61 ACAAAAAAGAAATAGCTGTCTTTTATCCAGGAAGGGGTAATAAGATAGAGTGGGAGATGG 120
Db     996 ACAAAAAAGAAATAGCTGTCTTTTATCCAGGAAGGGATAATAAGATAGAGTGGGAGATGG 1055

Qy     121 GCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCTAC 180
Db    1056 GCGCGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCTAC 1115

Qy     181 GACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTAG 240
Db    1116 GACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTAG 1175

Qy     241 ATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTTT 300
Db    1176 ATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTTT 1235

Qy     301 CGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACTG 360
Db    1236 CGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACTG 1295

Qy     361 TCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAAC 420
Db    1296 TCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAAC 1355

Qy     421 AGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAAACAA 480
Db    1356 AGATAGTACAGAGACACCTAGTGGTGGAAACAGGAACAGCAGAAATTATGCCAAAAACAA 1415

Qy     481 GTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGGTG 540
Db    1416 GCAGACCAACTGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGGTG 1475

Qy     541 GTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGATAG 600
Db    1476 GTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGATAG 1535

Qy     601 AGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGTTGCA 660
Db    1536 AGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGTTGCA 1595

Qy     661 CCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCAGA 720
Db    1596 CCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCAGA 1655

Qy     721 TTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACCAG 780
Db    1656 TTATCAGAGACATCATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACCAG 1715

Qy     781 CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACTAGTT 840
Db    1716 CTCCACAACAAGGACAGCTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACTAGTT 1775

Qy     841 CAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGCAACA 900
Db    1776 CAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGCAACA 1835

Qy     901 TTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCAACAA 960
Db    1836 TTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCAACAA 1895

Qy     961 ACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCT 1020
Db    1896 ACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCT 1955

Qy    1021 ACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACAC 1080
Db    1956 ACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACAC 2015

Qy    1081 TGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATC 1140
Db    2016 TGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATC 2075

Qy    1141 CCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCTA 1200
Db    2076 CCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCAGGACAGAAGGCTA 2135

Qy    1201 GATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCAGCAG 1260
Db    2136 GATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCAGCAG 2195

Qy    1261 CCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGACACT 1320
Db    2196 CCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGACACT 2255

Qy    1321 CTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATGGACC 1380
Db    2256 CTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATGGACC 2315

Qy    1381 ATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGGGGAA 1440
Db    2316 ATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGGGGAA 2375

Qy    1441 AGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGCTCCCC 1500
Db    2376 AGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGACGCCAACTGCTCCCC 2435

Qy    1501 CAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAGAGAG 1560
Db    2436 CAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAGAGAG 2495

Qy    1561 AAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCACC 1620
Db    2496 AAAAACAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGATGACAGAGGATTTGCTGCACC 2555

Qy    1621 TCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAGA 1680
Db    2556 TCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAGA 2615

Qy    1681 AGTATTACTGGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCACA 1740
Db    2616 AGTATTACTGGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCACA 2675

Qy    1741 TTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAAA 1800
Db    2676 TTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAAA 2735

Qy    1801 TGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCCC 1860
Db    2736 TGTAGAAATAGAAGTTTTAGGCAAAAGAATTAGAGGGACAATCATGACAGGGGACACCCC 2795

Qy    1861 GATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAAT 1912
Db    2796 GATTAACATTTTTGGTAGAAATTTACTAACAGCTCTGGGGATGTCTCTAAAT 2847
```

RESULT 10
SIVMM32H
LOCUS SIVMM32H 10277 bp DNA linear VRL 14-APR-2000
DEFINITION Simian immunodeficiency virus proviral DNA for env, gag, nef, pol, rev, tat, vif, vpr, and vpx proteins, complete cds.
ACCESSION D01065
VERSION D01065.1 GI:222602
KEYWORDS env polyprotein; gag polyprotein; nef protein; pol polyprotein; rev protein; tat protein; vif protein; vpr protein; vpx protein.
SOURCE Simian immunodeficiency virus (isolate:SIVmac32H, specific_host:Macaca mulatta) proviral DNA, clone:pJ5.

```
  ORGANISM  Simian immunodeficiency virus
            Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
            lentivirus group.
REFERENCE   1  (sites)
  AUTHORS   Rud,E.W., Yon,J.R., Larder,B.A., Cook,N., Cranage,M.P. and
            Clarke,B.E.
  TITLE     Infectious molecular clones of SIVmac32H: Nef deletion controls
            ability to reisolate virus from rhesus macaques
  JOURNAL   In Vaccines 92 (1992) In press
REFERENCE   2  (bases 1 to 10277)
  AUTHORS   Rud,E.W., Cranage,M., Yon,J., Quirk,J., Ogilvie,L., Cook,N.,
            Webster,S., Dennis,M. and Clarke,B.E.
  TITLE     Molecular and biological characterization of simian
            immunodeficiency virus macaque strain 32H proviral clones
            containing nef size variants
  JOURNAL   J. Gen. Virol. 75 (Pt 3), 529-543 (1994)
  MEDLINE   94172326
   PUBMED   8126450
COMMENT     Submitted (10-JUN-1991) to DDBJ by:
            Erling W Rud
            Molecular Sciences
            Wellcome Research Laboratories
            Langley Court, South Eden Park Road
            Beckenham, Kent BR3 3BS
            U.K.
            Phone:  081-658-2211 x5279
            Fax:    081-663-3532
            SIVmac32H proviral clone pJ5 contains no premature stop codons in
            any of the known SIV open reading frames. The variant pC8 is
            identical to pJ5 except for the changes indicated in the feature
            table.
FEATURES             Location/Qualifiers
     source          1. .10277
                     /organism="Simian immunodeficiency virus"
                     /proviral
                     /isolate="SIVmac32H"
                     /specific_host="Macaca mulatta"
                     /db_xref="taxon:11723"
                     /clone="pJ5"
                     /note="propagated in human T-cell (cell line JM)"
     LTR             1. .818
                     /note="5'LTR"
     LTR             1. .518
                     /note="5'LTR U3 region"
     protein_bind    415. .424
                     /bound_moiety="NFkB"
     protein_bind    429. .438
                     /bound_moiety="Sp1"
     protein_bind    440. .449
                     /bound_moiety="Sp1"
     protein_bind    451. .460
                     /bound_moiety="Sp1"
     protein_bind    462. .471
                     /bound_moiety="Sp1"
     TATA_signal     488. .494
     misc_feature    518. .519
                     /note="put. mRNA start site"
     repeat_unit     519. .694
                     /rpt_type=terminal
     LTR             695. .818
                     /note="5'LTR U5 region"
     primer_bind     822. .849
                     /note="tRNA (Lys) primer binding site"
     gene            1054. .2586
                     /gene="gag"
     CDS             1054. .2586
                     /gene="gag"
                     /codon_start=1
                     /product="gag polyprotein"
                     /protein_id="BAA00862.1"
                     /db_xref="GI:222607"
                     /translation="MGARNSVLSGKKADELEKIRLRPGGKKKYMLKHVVWAANELDRF
                     GLAESLLENKEGCQKILSVLAPLVPTGSENLKSLYNTVCVIWCIHAEEKVKHTEEAKQ
                     IVQRHLVVETGTAETMPKTSRPTAPSSGRGGNYPVQQIGGNYVHLPLSPRTLNAWVKL
                     IEEKKFGAEVVPGFQALSEGCTPYDINQMLNCVGDHQAAMQIIRDIINEEAADWDLQH
                     PQPAPQQGQLREPSGSDIAGTTSSVDEQIQWMYRQQNPIPVGNIYRRWIQLGLQKCVR
                     MYNPTNILDVKQGPKEPFQSYVDRFYKSLRAEQTDAAVKNWMTQTLLIQNANPDCKLV
                     LKGLGVNPTLEEMLTACQGVGGPGQKARLMAEALKEALAPVPIPFAAAQQRGPRKPIK
                     CWNCGKEGHSARQCRAPRRQGCWKCGKMDHVMAKCPDRQAGFLGLGPWGKKPRNFPMA
                     QVHQGLTPTAPPEDPAVDLLKNYMQLGKQQREKQRESREKPYKEVTEDLLHLNSLFGG
                     DQ"
     gene            2229. .5411
                     /gene="pol"
     CDS             <2229. .5411
                     /gene="pol"
                     /function="protease, reverse transcriptase, endonuclease"
                     /codon_start=1
                     /product="pol polyprotein"
                     /protein_id="BAA00863.1"
                     /db_xref="GI:2160378"
                     /translation="VLELWERRTLCKAMQSPKKTGMLEMWKNGPCYGQMPRQTGGFFR
                     PWSMGKEAPQFPHGSSASGADANCSPRGPSCGSAKELHAAGQAAERKAERKQREALQG
                     GDRGFAAPQFSLWRRPVVTAHIEGQPVEVLLDTGADDSIVTGIELGLHYTPKIVGGIG
                     GFINTKEYKNVEIEVLGKRIKGTIMTGDTPINIFGRNLLTALGMSLNLPIAKVEPVKV
                     ALKPGKDGPKLKQWPLSKEKIVALREICEKMEKDGQLEEAPPTNPYNTPTFAIKKKDK
                     NKWRMLIDFRELNRVTQDFTEVQLGIPHPAGLAKRKRITVLDIGDAYFSIPLDEEFRQ
                     YTAFTLPSVNNAEPGKRYIYKVLPQGWKGSPAIFQYTMRHVLEPFRKANPDVTLVQYM
                     DDILIASDRTDLEHDRVVLQLKELLNSIGFSTPEEKFQKDPPFQWMGYELWPTKWKLQ
                     KIELPQRETWTVNDIQKLVGVLNWAAQIYPGIKTKHLCRLIRGKMTLTEEVQWTEMAE
                     AEYEENKIILSQEQEGCYYQEGKPLEATVIKSQDNQWSYKIHQEDKILKVGKFAKIKN
                     THTNGVRLLAHVIQKIGKEAIVIWGQVPKFHLPVEKDVWEQWWTDYWQVTWIPEWDFI
                     STPPLVRLVFNLVKDPIKGEETYYTDGSCNKQSKEGKAGYITDRGKDKVKVLEQTTNQ
                     QAELEAFLMALTDSGPKANIIVDSQYVMGIITGCPTESESRLVNQIIEEMIKKSEIYV
                     AWVPAHKGIGGNQEIDHLVSQGIRQVLFLEKIEPAQEEHDKYHSNVKELVFKFGLPRI
                     VARQIVDTCDKCHQKGEAIHGQVNSDLGTWQMDCTHLEGKIIIVAVHVASGFIEAEVI
                     PQETGRQTALFLLKLAGRWPVTHLHTDNGANFASQEVKMVAWWAGIEHTFGVPYNPQS
                     QGVVEAMNHHLKNQIDRIREQANSVETIVLMAVHCMNFKRRGGIGDMTPAERLINMIT
                     TEQEIQFQQSKNSKFKNFRVYYREGRDQLWKGPGELLWKGEGAVILKVGTDIKVVPRR
                     KAKIIKDYGGGKEVDSSSHMEDTGEAREVA"
     repeat_region   2489. .2522
                     /rpt_type=direct
                     /rpt_unit=2489. .2500
                     /rpt_unit=2501. .2512
                     /rpt_unit=2513. .1522
     gene            5341. .5985
                     /gene="vif"
     CDS             5341. .5985
                     /gene="vif"
                     /function="virion infectivity factor"
                     /codon_start=1
                     /product="vif protein"
                     /protein_id="BAA00864.1"
                     /db_xref="GI:222609"
                     /translation="MEEGKRWIAVPTWRIPERLERWHSLIKYLKYKTKDLQKVCYVPH
                     YKVGWAWWTCSRVIFPLQEGSHLEVQGYWHLTPERGWLSTYAVRITWYSRNFWTDVTP
                     DYADILLHSTYFPCFTAGEVRRVIRGEQLLSCCKFPRAHRYQVPSLQYLALRVVSYVR
                     SQGENPTWKQWRRDSRRSLRMAKQNSRGDKQRGGKPPTKGANFPGLAKVLGILA"
     gene            5813. .6151
                     /gene="vpx"
     CDS             5813. .6151
                     /gene="vpx"
                     /codon_start=1
                     /product="vpx protein"
                     /protein_id="BAA00865.1"
                     /db_xref="GI:222610"
                     /translation="MSDPRERIPPGNSGEETVGEAFEWLNRTVEEINREAVNHLPREL
                     IFQVWQRSWEYWHDEQGMSQSYVKYRYLCLIQKALFMHCKKGCRCLGEGHGAGGWRPG
                     PPPPPPPGLA"
     gene            6152. .6457
                     /gene="vpr"
     CDS             6152. .6457
                     /gene="vpr"
                     /codon_start=1
                     /product="vpr protein"
                     /protein_id="BAA00866.1"
                     /db_xref="GI:222611"
```

```
                /translation="MEERPPENEGPQREPWDEWVVEVLEELKEEALKHFDPRLLTALG
                NHIYNRHGDTLEGAGELIRILQRALFMHFRGGCNHSRIGQPGGGNPLSTIPPSRGVL"
     gene       join(6303. .6598,8801. .8900)
                /gene="tat"
     CDS        join(6303. .6598,8801. .8900)
                /gene="tat"
                /function="transactivation"
                /codon_start=1
                /product="tat protein"
                /protein_id="BAA00867.1"
                /db_xref="GI:222603"
                /translation="METPLREQENSLESSNERSSCISEADATTPESANLGEEILSQLY
                RPLEACYNTCYCKKCCYHCQFCFLKKGLGICYEQSRKRRRTPKKAKANTSSASNNRPI
                SNRTRHCQPEKAKKETVEKAVATAPGLGR"
     CDS        join(6303. .6598,8804. .8900)
                /gene="tat"
                /note="alternative splicing"
                /codon_start=1
                /product="tat protein"
                /protein_id="BAA00868.1"
                /db_xref="GI:222605"
                /translation="METPLREQENSLESSNERSSCISEADATTPESANLGEEILSQLY
                RPLEACYNTCYCKKCCYHCQFCFLKKGLGICYEQSRKRRRTPKKAKANTSSASNKPIS
                NRTRHCQPEKAKKETVEKAVATAPGLGR"
     exon       <6303. .6598
                /gene="tat"
                /product="tat protein"
                /number=1
     gene       join(6529. .6598,8801. .9057)
                /gene="rev"
     CDS        join(6529. .6598,8801. .9057)
                /gene="rev"
                /codon_start=1
                /product="rev protein"
                /protein_id="BAA00869.1"
                /db_xref="GI:222604"
                /translation="MSSHEREEELRKRLRLIHLLHQTTDPYPTGPGTANQRRQRRRRW
                RRRWQQLLALADRIYSFPDPPTDTPLDLAIQQLQNLAIESIPDPPTNTPEALCGPTEN
                SRSPQD"
     CDS        join(6529. .6598,8804. .9057)
                /gene="rev"
                /note="alternative splicing"
                /codon_start=1
                /product="rev protein"
                /protein_id="BAA00870.1"

  Query Match             97.2%;  Score 1860;  DB 14;  Length 10277;
  Best Local Similarity   98.6%;  Pred. No. 0;
  Matches 1886;  Conservative    0;  Mismatches   25;  Indels    1;  Gaps    1;

Qy      1 GGTACCAGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGCAAC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    937 GGTACCAGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGCAAC 996

Qy     61 AC-AAAAAAGAAATAGCTGTCTTTTATCCAGGAAGGGGTAATAAGATAGAGTGGGAGATG 119
          || |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    997 ACAAAAAAAGAAATAGCTGTCTTTTATCCAGGAAGGGATAATAAGATAGAGTGGGAGATG 1056

Qy    120 GGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCTA 179
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1057 GGCGCGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCTA 1116

Qy    180 CGACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTA 239
          ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1117 CGACCCGGCGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTA 1176

Qy    240 GATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTT 299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1177 GATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTT 1236

Qy    300 TCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACT 359
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1237 TCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACT 1296

Qy    360 GTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAA 419
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1297 GTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAA 1356

Qy    420 CAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAAACA 479
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   1357 CAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAGCAGAAACTATGCCAAAAACA 1416

Qy    480 AGTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGGT 539
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1417 AGTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGGT 1476

Qy    540 GGTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGATA 599
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1477 GGTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGATA 1536

Qy    600 GAGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGTTGC 659
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   1537 GAGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAAGCACTGTCAGAAGGCTGC 1596

Qy    660 ACCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCAG 719
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1597 ACCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCAG 1656

Qy    720 ATTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACCA 779
          |||||||||| |||||||| ||||||||||||||||||||||||||||||||||||||||
Db   1657 ATTATCAGAGACATTATAAATGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACCA 1716

Qy    780 GCTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACTAGT 839
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   1717 GCTCCACAACAAGGACAGCTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACTAGT 1776

Qy    840 TCAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGCAAC 899
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1777 TCAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGCAAC 1836

Qy    900 ATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCAACA 959
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1837 ATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCAACA 1896

Qy    960 AACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTC 1019
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   1897 AACATTCTAGATGTAAAACAAGGACCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTC 1956

Qy   1020 TACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACA 1079
          |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1957 TACAAAAGCTTAAGAGCAGAACAAACAGATGCAGCAGTAAAGAATTGGATGACTCAAACA 2016

Qy   1080 CTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAAT 1139
          |||||||||||||||||||||||||||||||| |||| |||||||| ||||||||||||
Db   2017 CTGCTGATTCAAAATGCTAACCCAGATTGCAAGTTAGTACTGAAGGGACTGGGTGTGAAT 2076

Qy   1140 CCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCT 1199
          |||||||||||||||||||||||||||||||||||||||||||| || ||||||||||||
Db   2077 CCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGACCAGGACAGAAGGCT 2136

Qy   1200 AGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCAGCA 1259
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2137 AGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCAGCA 2196

Qy   1260 GCCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGACAC 1319
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   2197 GCCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAAGGACAC 2256

Qy   1320 TCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATGGAC 1379
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2257 TCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATGGAC 2316

Qy   1380 CATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGGGGA 1439
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2317 CATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGGGGA 2376
```

```
Qy    1440 AAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGCTCCC 1499
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2377 AAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGACGCCAACTGCTCCC 2436

Qy    1500 CCAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAGAGA 1559
           ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    2437 CCAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGCTGGGCAAGCAGCAGAGA 2496

Qy    1560 GAAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCAC 1619
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2497 GAAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCAC 2556

Qy    1620 CTCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAG 1679
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2557 CTCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAG 2616

Qy    1680 AAGTATTACTGGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCAC 1739
           |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    2617 AAGTATTACTAGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCTAC 2676

Qy    1740 ATTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAA 1799
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2677 ATTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAA 2736

Qy    1800 ATGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCC 1859
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2737 ATGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCC 2796

Qy    1860 CGATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAA 1911
           |||||||||||||||||||||| |||||||||||||||||||| ||||
Db    2797 CGATTAACATTTTTGGTAGAAATTTACTAACAGCTCTGGGGATGTCTTTAAA 2848
```

RESULT 11
SIVMM251
LOCUS SIVMM251 10277 bp ss-RNA linear VRL 29-NOV-2000
DEFINITION Simian (macaque) immunodeficiency virus, complete genome; isolate Mm251, proviral DNA.
ACCESSION M19499 M15897 M16125 M24614 Y00294
VERSION M19499.1 GI:334657
KEYWORDS env protein; envelope-associated protein; gag polyprotein; nef protein; pol polyprotein; rev protein; tat protein; vif protein; vpr protein; vpx protein.
SOURCE Simian immunodeficiency virus.
ORGANISM Simian immunodeficiency virus
Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate lentivirus group.
REFERENCE 1 (bases 5731 to 10277)
AUTHORS Hirsch,V.M.
JOURNAL Unpublished
REFERENCE 2 (bases 809 to 1677; 5731 to 10249)
AUTHORS Hahn,B.H., Kumar,P., Taylor,M.E., Arya,S.K. and Shaw,G.M.
TITLE Human retrovirses, cancer and AIDS: approach to prevention and therapy
JOURNAL Unpublished
REFERENCE 3 (bases 6078 to 10249)
AUTHORS Hirsch,V., Riedel,N. and Mullins,J.I.
TITLE The genome organization of STLV-3 is similar to that of the AIDS virus except for a truncated transmembrane protein
JOURNAL Cell 49 (3), 307-319 (1987)
MEDLINE 87187627
PUBMED 3646094
REFERENCE 4 (bases 506 to 5975; 6319 to 10125)
AUTHORS Franchini,G., Gurgo,C., Guo,H.G., Gallo,R.C., Collalti,E., Fargnoli,K.A., Hall,L.F., Wong-Staal,F. and Reitz,M.S. Jr.
TITLE Sequence of simian immunodeficiency virus and its relationship to the human immunodeficiency viruses
JOURNAL Nature 328 (6130), 539-543 (1987)
MEDLINE 87287229
PUBMED 3497350
REFERENCE 5 (bases 7757 to 8090; 8108 to 8225; 8237 to 8625; 8706 to 8898)
AUTHORS Kestler,H.W.
JOURNAL Unpublished
REFERENCE 6 (bases 1 to 10249)
AUTHORS Donahue,P.R., Kornfeld,H., Gallo,M.V. and Mullins,J.I.
JOURNAL Unpublished
REFERENCE 7 (sites)
AUTHORS Kestler,H.W. III, Li,Y., Naidu,Y.M., Butler,C.V., Ochs,M.F., Jaenel,G., King,N.W., Daniel,M.D. and Desrosiers,R.C.
TITLE Comparison of simian immunodeficiency virus isolates
JOURNAL Nature 331 (6157), 619-622 (1988)
MEDLINE 88122665
PUBMED 2893293
COMMENT [7] sites.
Kestler et al. [7], (Nature 331, 619-622 (1988)) present strong evidence that the isolates previously referred to as STLV-III AGM ([4], [3], [1]) and HTLV-IV ([2]) are not authentic, but were derived from cell cultures infected with SIV MAC-251. The reference sequence for this entry is from unpublished (1988), with exception of bases
10250-10277 which are from [1].
the vpr coding region has been annotated to end near where the gene ends in the SIVMM142 sequence. A base change from 'a' at position 6443 in SIVMM142 to 't' at corresponding position 6418 in the SIVMM251 sequence results in the coding sequence ending 4 amino acids earlier than it does in SIVMM142. Also, the variation marked at position 6377 ('aa' in K6W, 'aga' in K78) implies different amino acid sequences for those isolates from that point onward in vpr.
The tat start is annotated in agreement with the site in [2] (also agrees with HIV2ROD) sequence rather than at the site marked in [3] for the K78 sequence.
The start codon for rev is marked at position 6504 in accordance with reference [3] for the K78 sequence. This position also agrees with the rev start in the SIVMM142 sequence. However, [2] annotates the rev start at position 6354. Both of these start sites are in the same frame. The sequence annotation shows two alternate 3' splice junction sites at 8743 and 8785. See reference [3]. Clean copy of sequence unpublished (1988) kindly provided by H.W. Kestler to the
EMBL data library (25-FEB-1988). Computer-readable copy of sequence unpublished (1988) kindly provided by J.Mullins (12-SEP-1988). An internal stop codon is present in the env cds at positions 8782-8784 in unpublished (1988) and 8785-8787 in [4],[3],[1], [2],unpublished (1988).
unpublished (1988) has shown proviral clone BK28 to be biologically active,
causing persistent lymphadenopathy and death in one out of four rhesus monkeys 17 months after infection. BK28 was originally called HTLV-4 (Nature 326, 610-613 (1987)), but it has since been found to correspond to SIVMM251. The K6W sequence [4], which was derived from isolate Mm251, contains about 25 frame shifts in gag-pol relative to this sequence.
A single base insertion at position 9678 results in a frame-shift in the nef coding region. The nef coding region in [4] would produce a protein about 36 residues longer than that produced by the other isolates.
Author [7] address Kestler H.W., Harvard Medical School, New England Regional Primate Research Center, Department of Microbiology, One Pine Hill Drive, Southborough, Mass. 01772, USA.
[7] submitted (25-FEB-1988) to the EMBL data library by. Complete source information:
Simian (macaque) immunodeficiency virus proviral DNA, SIV mac isolate K6W, clones lambda-16, 27, 35, and 11 [4]; and SIV mac isolate K78, clones K2-10 and K2-5 [3], [1]; HTLV4/PK82 and HTLV4/PK190 proviral DNA clones [2]; SIV mac integrated proviral DNA isolate 251, clone lambda 251 unpublished (1988); proviral DNA, isolate Mm251,
clone BK28 unpublished (1988).
FEATURES Location/Qualifiers
source 1. .10277
/organism="Simian immunodeficiency virus"
/proviral
/isolate="Mm251"
/db_xref="taxon:11723"
LTR 1. .806

```
                /note="5' LTR"
misc_binding    428. .437
                /note="putative"
                /bound_moiety="Sp1 III"
misc_binding    439. .448
                /note="putative"
                /bound_moiety="Sp1 II"
misc_binding    450. .459
                /note="putative"
                /bound_moiety="Sp1"
repeat_region   506. .682
                /note="R repeat 5' copy"
prim_transcript 507. .10125
                /note="genomic mRNA"
prim_transcript 507. .10125
                /note="tat, rev, nef subgenomic mRNA"
variation       517. .519
                /note="cgg in unpublished (1988); gc in [4]"
                /replace="gc"
variation       617. .618
                /note="tg in unpublished (1988); t in [4]"
                /replace="t"
variation       668
                /note="g in unpublished (1988); gc in [4]"
                /replace="gc"
variation       694
                /note="g in unpublished (1988); c in [4]"
                /replace="c"
variation       779
                /note="a in unpublished (1988); g in [4]"
                /replace="g"
variation       806
                /note="a in unpublished (1988); agt in [4]"
                /replace="agt"
misc_binding    810. .827
                /bound_moiety="Lys-tRNA primer"
variation       825. .826
                /note="ac in [4],unpublished (1988); gac in [2]"
                /replace="gac"
variation       843. .844
                /note="ga in [2],unpublished (1988); agaa in [4]"
                /replace="agaa"
variation       858. .859
                /note="ct in [2],unpublished (1988); cct in [4]"
                /replace="cct"
variation       870
                /note="a in [4],unpublished (1988); g in [Unpublished
                (1987) UCLA Symposia on Molecular and Cellular Biology]"
                /replace="g"
variation       895. .897
                /note="gac in [4],unpublished (1988); gc in [2]"
                /replace="gc"
variation       919
                /note="g in [2],unpublished (1988); a in [4]"
                /replace="a"
variation       966. .968
                /note="gtt in [4],unpublished (1988); gt in [2]"
                /replace="gt"
variation       1009. .1011
                /note="atc in [2],unpublished (1988); ac in [4]"
                /replace="ac"
variation       1014
                /note="g in [4],unpublished (1988); a in [2]"
                /replace="a"
CDS             1041. .2561
                /note="gag polyprotein"
                /codon_start=1
                /protein_id="AAB59905.1"
                /db_xref="GI:334660"
                /translation="MGARNSVLSGKKADELEKIRLRPGGKKKYMLKHVVWAANELDRF
                GLAESLLENKEGCQKILSVLAPLVPTGSENLKSLYNTVCVIWCIHAEEKVKHTEEAKQ
                IVQRHLVVETGTAETMPKTSRPTAPSSGRGGNYPVQQIGGNYVHLPLSPRTLNAWVKL
                IEEKKFGAEVVPGFQALSEGCTPYDINQMLNCVGDHQAAMQIIRDIINEEAADWDLQH
                PQPAPQQGQLREPSGSDIAGTTSSVDEQIQWMYRQQNPIPVGNIYRRWIQLGLQKCVR
                MYNPTNILDVKQGPKEPFQSYVDRFYKSLRAEQTDAAVKNWMTQTLLIQNANPDCKLV
                LKGLGVNPTLEEMLTACQGVGGPGQKARLMAEALKEALAPVPIPFAAAQKRGPRKPIK
                CWNCGKEGHSARQCRAPRRQGCWKCGKMDHVMAKCPDRQAGFLGLGPWGKKPRNFPMA
                QVHQGLTPTAPPEDPAVDLLKNYMQLGKQQRESREKPYKEVTEDLLHLNSLFGGDQ"
variation       1056
                /note="t in [2],unpublished (1988); g in [4]"
                /replace="g"
variation       1083
                /note="g in [4],unpublished (1988); a in [2]"
                /replace="a"
variation       1190
                /note="g in [4],unpublished (1988); c in [4]"
                /replace="c"
variation       1365
                /note="g in [4],unpublished (1988); a in [2]"
```

Query Match 95.8%; Score 1833.4; DB 14; Length 10277;
Best Local Similarity 98.2%; Pred. No. 0;
Matches 1879; Conservative 0; Mismatches 21; Indels 13; Gaps 2;

```
Qy      1 GGTACCAGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGCAAC 60
          ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db    924 GGTACCAGACGGCGTGAGGAGCGGGAGAGGAGGAGGCCTCCGGTTGCAGGTAAGTGCAAC 983

Qy     61 ACAAAAAAGAAATAGCTGTCTT-TTATCCAGGAAGGGGTAATAAGATAGAGTGGGAGATG 119
          |||||||||||||||||||||| |||||||||||||| ||||||||||||||||||||||
Db    984 ACAAAAAAGAAATAGCTGTCTTGTTATCCAGGAAGGGATAATAAGATAGAGTGGGAGATG 1043

Qy    120 GGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCTA 179
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1044 GGCGCGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCTA 1103

Qy    180 CGACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTA 239
          ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1104 CGACCCGGCGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTA 1163

Qy    240 GATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTT 299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1164 GATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTT 1223

Qy    300 TCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACT 359
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1224 TCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACT 1283

Qy    360 GTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAA 419
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1284 GTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAA 1343

Qy    420 CAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAAACA 479
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   1344 CAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAGCAGAAACTATGCCAAAAACA 1403

Qy    480 AGTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGGT 539
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1404 AGTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGGT 1463

Qy    540 GGTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGATA 599
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1464 GGTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGATA 1523

Qy    600 GAGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGTTGC 659
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   1524 GAGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGCTGC 1583

Qy    660 ACCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCAG 719
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1584 ACCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCAG 1643

Qy    720 ATTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACCA 779
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   1644 ATTATCAGAGATATTATAAATGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACCA 1703
```

```
Qy    780 GCTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACTAGT 839
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   1704 GCTCCACAACAAGGACAGCTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACTAGT 1763

Qy    840 TCAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGCAAC 899
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1764 TCAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGCAAC 1823

Qy    900 ATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCAACA 959
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1824 ATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCAACA 1883

Qy    960 AACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTC 1019
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1884 AACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTC 1943

Qy   1020 TACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACA 1079
          |||||||| |||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   1944 TACAAAAGCTTAAGAGCAGAACAAACAGATGCAGCAGTAAAGAATTGGATGACTCAAACA 2003

Qy   1080 CTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAAT 1139
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2004 CTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAAT 2063

Qy   1140 CCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCT 1199
          |||||||||||||||||||||||||||||||||||| || ||||||||||||
Db   2064 CCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGACCAGGACAGAAGGCT 2123

Qy   1200 AGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCAGCA 1259
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2124 AGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCAGCA 2183

Qy   1260 GCCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGACAC 1319
          |||||  ||||||||||||||||||||||||||||||||||||||| ||||||||
Db   2184 GCCCAGAAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAGGAGGGACAC 2243

Qy   1320 TCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATGGAC 1379
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2244 TCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATGGAC 2303

Qy   1380 CATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGGGGA 1439
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2304 CATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGGGGA 2363

Qy   1440 AAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGCTCCC 1499
          |||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db   2364 AAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGACGCCAACTGCTCCC 2423

Qy   1500 CCAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAGAGA 1559
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2424 CCAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGC-------- 2475

Qy   1560 GAAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCAC 1619
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2476 ----AGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCAC 2531

Qy   1620 CTCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAG 1679
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2532 CTCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAG 2591

Qy   1680 AAGTATTACTGGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCAC 1739
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2592 AAGTATTATTGGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCAC 2651

Qy   1740 ATTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAA 1799
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2652 ATTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAA 2711

Qy   1800 ATGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCC 1859
          ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db   2712 ATGTAAAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACTC 2771

Qy   1860 CGATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAAT 1912
          |||||||||||||||| |||||||||||||||||||||||||||||||||||
Db   2772 CGATTAACATTTTTGGTAGGAATTTGCTAACAGCTCTGGGGATGTCTCTAAAT 2824
```

```
RESULT 12
RESIVMXX
LOCUS       RESIVMXX                9646 bp    DNA     linear   VRL 01-JUL-1999
DEFINITION  Simian immunodeficiency virus STLV-III(MAC) proviral genome.
ACCESSION   Y00277 M16403
VERSION     Y00277.1  GI:61730
KEYWORDS    art gene; env gene; envelope protein; gag gene; genome; overlapping
            genes; pol gene; tat gene.
SOURCE      Simian immunodeficiency virus.
  ORGANISM  Simian immunodeficiency virus
            Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
            lentivirus group.
REFERENCE   1  (bases 1 to 9646)
  AUTHORS   Chakrabarti,L., Guyader,M., Alizon,M., Daniel,M.D.,
            Desrosiers,R.C., Tiollais,P. and Sonigo,P.
  TITLE     Sequence of simian immunodeficiency virus from macaque and its
            relationship to other human and simian retroviruses
  JOURNAL   Nature 328 (6130), 543-547 (1987)
  MEDLINE   87287230
REFERENCE   2  (bases 1 to 9646)
  AUTHORS   Chakrabarti,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-AUG-1987)
COMMENT     On Jul 24, 2002 this sequence version replaced gi:334638.
FEATURES             Location/Qualifiers
     source          1. .9646
                     /organism="Simian immunodeficiency virus"
                     /db_xref="taxon:11723"
                     /clone="lambda SIV1"
                     /clone_lib="infected HUT78 cells"
     repeat_region   1. .318
                     /note="5' LTR"
     misc_feature    1. .175
                     /note="LTR (R)"
     misc_feature    176. .318
                     /note="LTR (U5)"
     misc_feature    488. .2068
                     /note="ORF (gag)"
     CDS             551. .2071
                     /note="gag gene product (AA 1-506)"
                     /codon_start=1
                     /protein_id="CAA68379.1"
                     /db_xref="GI:61731"
                     /db_xref="SWISS-PROT:P05894"
                     /translation="MGARNSVLSGKKADELEKIRLRPGGKKKYMLKHVVWAANELDRF
                     GLAESLLENKEGCQKILSVLAPLVPTGSENLKSLYNTVCVIWCIHAEEKVKHTEEAKQ
                     IVQRHLVMETGTAETMPKTSRPTAPFSGRGGNYPVQQIGGNYTHLPLSPRTLNAWVKL
                     IEEKKFGAEVVSGFQALSEGCLPYDINQMLNCVGDHQAAMQIIRDIINEEAADWDLQH
                     PQQAPQQGQLREPSGSDIAGTTSTVEEQIQWMYRQQNPIPVGNIYRRWIQLGLQKCVR
                     MYNPTNILDVKQGPKEPFQSYVDRFYKSLRAEQTDPAVKNWMTQTLLIQNANPDCKLV
                     LKGLGTNPTLEEMLTACQGVGGPGQKARLMAEALKEALAPAPIPFAAAQQKGPRKPIK
                     CWNCGKEGHSARQCRAPRRQGCWKCGKMDHVMAKCPNRQAGFLGLGPWGKKPRNFPMA
                     QVHQGLTPTAPPEEPAVDLLKNYMHLGKQQRESRGKPYKEVTEDLLHLNSLFGGDQ"
     CDS             <1726. .4896
                     /note="pol gene product (AA 1-1056)"
                     /codon_start=1
                     /protein_id="CAA68380.1"
                     /db_xref="GI:1335590"
                     /db_xref="SWISS-PROT:P05896"
                     /translation="VLELWEGRTLCKAMQSPKKTGMLEMWKNGPCYGQMPKQTGGFFR
                     PWPLGKEAPQFPHGSSASGADANCSPRRTSCGSAKELHALGQAAERKQREALQGGDRG
                     FAAPQFSLWRRPVVTAHIEGQPVEVLLDTGADDSIVTGIELGPHYTPKIVGGIGGFIN
                     TKEYKNVEIEVLGKRIKGTIMTGDTPINIFGRNLLTALGMSLNLPIAKVEPVKSPLKP
                     GKDGPKLKQWPLSKEKIVALREICEKMEKDGQLEEAPPTNPYNTPTFAIKKKDKNKWR
                     MLIDFRELNRVTQDFTEVQLGIPHPAGLAKRKRITVLDIGDAYFSIPLDEEFRQYTAF
                     TLPSVNNAEPGKRYIYKVLPQGWKGSPAIFQYTMRHVLEPFRKANPDVTLVQYMDDIL
                     IASDRTDLEHDRVVLQLKELLNSIGFSSPEEKFQKDPPFQWMGYELWPTKWKLQKIEL
                     PQRETWTVNDIQKLVGVLNWAAQIYPGIKTKHLCRLIRGKMTLTEEVQWTEMAEAEYE
                     ENKIILSQEQEGCYYQESKPLEATVIKSQDNQWSYKIHQEDKILKVGKFAKIKNTHTN
```

```
                     GVRLLAHVIQKIGKEAIVIWGQVPKFHLPVEKDVWEQWWTDYWQVTWIPEWDFISTPP
                     LVRLVFNLVKDPIEGEETYYVDGSCSKQSKEGKAGYITDRGKDKVKVLEQTTNQQAEL
                     EAFLMALTDSGPKANIIVDSQYVMGIITGCPTESESRLVNQIIEEMIKKTEIYVAWVP
                     AHKGIGGNQEIDHLVSQGIRQVLFLEKIEPAQEEHSKYHSNIKELVFKFGLPRLVAKQ
                     IVDTCDKCHQKGEAIHGQVNSDLGTWQMDCTHLEGKIVIVAVHVASGFIEAEVIPQET
                     GRQTALFLLKLASRWPITHLHTDNGANFASQEVKMVAWWAGIEHTFGVPYNPQSQGVV
                     EAMNHHLKNQIDRIREQANSVETIVLMAVHCMNFKRRGGIGDMTPAERLINMITTEQE
                     IQFQQSKNSKFKNFRVYYREGRDQLWKGPGELLWKGEGAVILKVGTDIKVVPRRKAKI
                     IKDYGGGKEMDSSSHMEDTGEAREVA"
misc_feature         4793. .5467
                     /note="ORF (Q)"
CDS                  4826. .5470
                     /note="Q gene product (AA 1-214)"
                     /codon_start=1
                     /protein_id="CAA68381.1"
                     /db_xref="GI:61733"
                     /db_xref="SWISS-PROT:P05903"
                     /translation="MEEEKRWIVVPTWRIPERLERWHSLIKYLKYKTKDLQKACYVPH
                     HKVGWAWWTCSRVIFPLQEGSHLEVQGYWNLTPERGWLSTYAVRITWYSKDFWTDVTP
                     EYADILLHSTYFPCFTAGEVRRAIRGERLLSCCRFPRAHKHQVPSLQYLALRVVSHVR
                     SQGENPTWKQWRRDNRRSLRVAKQNSRGDKQRGGKPPTEGANFPGLAKVLGILA"
misc_feature         5295. .5633
                     /note="ORF (X)"
CDS                  5298. .5636
                     /note="X gene product (AA 1-112)"
                     /codon_start=1
                     /protein_id="CAA68382.1"
                     /db_xref="GI:61734"
                     /db_xref="SWISS-PROT:P05917"
                     /translation="MSDPRERIPPGNSGEETIGEAFEWLNRTVEEINREAVNHLPREL
                     IFQVWQRSWEYWHDEQGMSQSYTKYRYLCLIQKALFMHCKKGCRCLGEGHGAGGWRPG
                     PPPPPPPGLA"
CDS                  5637. .5942
                     /note="R gene product (AA 1-101)"
                     /codon_start=1
                     /protein_id="CAA68383.1"
                     /db_xref="GI:61735"
                     /db_xref="SWISS-PROT:P05958"
                     /translation="MEERPPENEGPQREPWDEWVVEVLKELKEEALKHFDPRLLTALG
                     NHIYNRHGDTLEGAGELIRILQRALFIHFRSGCSHSRIGQPGGGNPLSTIPPSRSML"
misc_feature         5758. .6083
                     /note="ORF, exon 1 (tat)"
CDS                  join(5788. .6083,8301. .8397)
                     /note="tat gene product"
                     /codon_start=1
                     /protein_id="CAB46521.1"
                     /db_xref="GI:5419939"
                     /translation="METPLREQENSLESSNERSSYISEAAAAIPESANLGEEILSQLY
                     RPLEACYNTCYCKKCCYHCQFCFLKKGLGISYEKSHRRRRTPKKAKANTSSASNEPIP
                     NRIRLCQPKKAKKETVEAAVATAPGLGR"
misc_feature         6011. .6083
                     /note="ORF, exon 1 (art)"
CDS                  join(6014. .6083,8301. .8554)
                     /note="art gene product"
                     /codon_start=1
                     /protein_id="CAB46522.1"
                     /db_xref="GI:5419940"
                     /translation="MRSHTGEEELRRRLRLIHLLHQTNPYPTGSGSANQRRQKRRRWR
                     QRWQQLLALADRIYSFPDPPTDTPLDLAIQQLQNLAIESIPDPPTNIPEALCDPTENS
                     RSPQA"
misc_feature         6067. .8735
                     /note="ORF (env)"
intron               6084. .8300
                     /note="intron I (tat and art) (8301 is 3rd base in codon)"
mRNA                 6090. .8735
                     /note="envelope protein (AA 1-882)"
misc_feature         8247. .8551
                     /note="ORF, exon 2 (art)"
misc_feature         8251. .8394
                     /note="ORF, exon 2 (tat)"
misc_feature         8298. .8300
                     /note="in-frame stop codon (env)"
misc_feature         8560. .9357
                     /note="ORF (F)"
CDS                  8572. .9360
                     /note="F-gene product (AA 1-262)"
                     /codon_start=1
                     /protein_id="CAA68389.1"
                     /db_xref="GI:61740"
                     /db_xref="SWISS-PROT:P05862"
                     /translation="MGGAISKKRSKPPRDLRQRLLRARGENYGRLFKGVEDGSSQSLG
                     GLDKGLSSLSCEGQKYNQGEYMNTPWRNPAEERKKLPYRKQNIDDIEEDDDLVGIPV
                     EARVPLRTMSYKLAIDMSHFIKEKGGLEGIYYSARRHRILDIYLEKEEGIIPDWQIHS
                     GPGIRYLKMFGWLWKLIPVNVSDEAQEDEEHYLVHPAQTSQWDDPWGEVLAWKFDPTL
                     AYTYEAYIRYPEEFGSKSGLSEKEVKRRLAARGLLEMADRKETS"
repeat_region        8957. .9646
                     /note="3' LTR"
misc_feature         8957. .9471
                     /note="LTR (U3)"
promoter             9368. .9377
                     /note="enhancer-like sequence"
promoter             9382. .9391
                     /note="enhancer-like sequence"
misc_feature         9392. .9402
                     /note="conserved sequence"
misc_feature         9404. .9412
                     /note="conserved sequence"
misc_feature         9414. .9424
                     /note="conserved sequence"
misc_feature         9472. .9646
                     /note="LTR (R)"
promoter             9491. .9496
                     /note="pot. promoter sequence"
BASE COUNT     3290 a   1817 c   2387 g   2152 t
ORIGIN

  Query Match           93.5%;  Score 1788.6;  DB 14;  Length 9646;
  Best Local Similarity   96.8%;  Pred. No. 0;
  Matches 1851;  Conservative    0;  Mismatches   49;  Indels   13;  Gaps    2;

Qy       1 GGTACCAGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCC-GGTTGCAGGTAAGTGCAA 59
           ||||||||||||||||||||||||||    |||||||||||| ||||||||||||||||||
Db     434 GGTACCAGACGGCGTGAGGAGCGGGGAGAGAAGAGGCCTCCTGGTTGCAGGTAAGTGCAA 493

Qy      60 CACAAAAAAGAAATAGCTGTCTTTTATCCAGGAAGGGGTAATAAGATAGAGTGGGAGATG 119
           |||||||| ||||||||||||||||||||||||||| |||||||||||||||||||||||
Db     494 CACAAAAAGGAAATAGCTGTCTTTTATCCAGGAAGGGATAATAAGATAGAGTGGGAGATG 553

Qy     120 GGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCTA 179
           |||| |||||||||||||||||||||||||||||||||||||||||||||||| ||| |||
Db     554 GGCGCGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGACTA 613

Qy     180 CGACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTA 239
           ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     614 CGACCCGGCGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTA 673

Qy     240 GATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTT 299
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     674 GATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTT 733

Qy     300 TCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACT 359
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     734 TCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACT 793

Qy     360 GTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAA 419
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     794 GTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAA 853

Qy     420 CAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAAACA 479
           |||||||||||||||||||||| |||||||||||||| |||||||||||||||||||||
Db     854 CAGATAGTGCAGAGACACCTAGTGATGGAAACAGGAACAGCAGAAACTATGCCAAAAACA 913

Qy     480 AGTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGGT 539
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     914 AGTAGACCAACAGCACCATTTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGGT 973
```

```
Qy    540 GGTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGATA 599
          ||||||||||  ||||||  ||||||||||||||||||||||||||||||||||||||||
Db    974 GGTAACTATACCCACCTACCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTAATA 1033

Qy    600 GAGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGTTGC 659
          |||||  ||||||||||||||||||||||||  |||||||||||||||||||||||||| |||
Db   1034 GAGGAGAAGAAATTTGGAGCAGAAGTAGTGTCAGGATTTCAGGCACTGTCAGAAGGCTGC 1093

Qy    660 ACCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCAG 719
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1094 CTCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCAG 1153

Qy    720 ATTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACCA 779
          ||  ||||||||||||||||  ||||||||||||||||||||||||||||||||||| |
Db   1154 ATCATCAGAGATATTATAAATGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACAA 1213

Qy    780 GCTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACTAGT 839
          ||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||
Db   1214 GCTCCACAACAAGGACAGCTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACTAGT 1273

Qy    840 TCAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGCAAC 899
          |||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1274 ACAGTAGAAGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGCAAC 1333

Qy    900 ATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCAACA 959
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1334 ATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCAACA 1393

Qy    960 AACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTC 1019
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1394 AACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTC 1453

Qy   1020 TACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACA 1079
          ||||||||||||||||||||||  ||||||  ||||||||||||||||||||||||||||
Db   1454 TACAAAAGTTTAAGAGCAGAACAAACAGATCCAGCAGTAAAGAATTGGATGACTCAAACA 1513

Qy   1080 CTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAAT 1139
          ||||||||||||||||||||||||||||||||||||||||||||||||||||  ||||
Db   1514 CTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTACGAAT 1573

Qy   1140 CCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCT 1199
          ||||||||||||||||||||||||  ||||||||||||||||||  ||||||||||||
Db   1574 CCCACCCTAGAAGAAATGCTGACGGCCTGTCAAGGAGTAGGGGGGCCAGGACAGAAGGCT 1633

Qy   1200 AGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCAGCA 1259
          ||||||||||||||||||||||||||||||||||||  ||||||||||||||||||||
Db   1634 AGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGCGCCAATCCCTTTTGCAGCA 1693

Qy   1260 GCCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGACAC 1319
          |||||||||  ||||||||||||||||||||||||||||||||||||||  || ||||||
Db   1694 GCCCAACAGAAGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAGGAAGGACAC 1753

Qy   1320 TCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATGGAC 1379
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1754 TCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATGGAC 1813

Qy   1380 CATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGGGGA 1439
          ||||||||||||||||||  ||||||||||||||||||||||||||| || ||||||
Db   1814 CATGTTATGGCCAAATGCCCAAACAGACAGGCGGGTTTTTTAGGCCTTGGCCCTTGGGGA 1873

Qy   1440 AAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGCTCCC 1499
          ||||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||
Db   1874 AAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGACGCCAACTGCTCCC 1933

Qy   1500 CCAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAGAGA 1559
          |||||  ||  ||||||||||||||||||||||||||||||||  ||||||||||
Db   1934 CCAGAAGAACCAGCTGTGGATCTGCTAAAGAACTACATGCACTTGGGCAAGC-------- 1985

Qy   1560 GAAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCAC 1619
          ||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||
Db   1986 ----AGCAGAGAGAAAGCAGAGGGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCAC 2041

Qy   1620 CTCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAG 1679
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2042 CTCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAG 2101

Qy   1680 AAGTATTACTGGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCAC 1739
          ||||||||  |  ||||||||||||||||||||||||||||||||||||||||||||||
Db   2102 AAGTATTATTAGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCAC 2161

Qy   1740 ATTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAA 1799
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2162 ATTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAA 2221

Qy   1800 ATGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCC 1859
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2222 ATGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCC 2281

Qy   1860 CGATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAAT 1912
          |||||||||||||||||||||||  ||||||||||||||||||||||||||
Db   2282 CGATTAACATTTTTGGTAGAAATTTACTAACAGCTCTGGGGATGTCTCTAAAT 2334
```

```
RESULT 13
SIVMNE
LOCUS       SIVMNE                  9628 bp ss-RNA     linear   VRL 05-MAR-1996
DEFINITION  Simian immunodeficiency virus (SIV) complete proviral genome.
ACCESSION   M32741
VERSION     M32741.1  GI:334692
KEYWORDS    complete genome.
SOURCE      Simian immunodeficiency virus (clone:  8.) cDNA to genomic RNA.
  ORGANISM  Simian immunodeficiency virus
            Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
            lentivirus group.
REFERENCE   1  (bases 1 to 9628)
  AUTHORS   Benveniste,R.E., Heidecker,G., Greenwood,J. and Gonda,M.A.
  JOURNAL   Unpublished (1990)
COMMENT     Kindly submitted in computer-readable form by R. Benveniste.  The
            gag protein sequence was reported in J. Virol. 62, 2587-2595, 1988.
            This molecular clone, after transfection into T-cell lines,
            produces infectious viral particles. In particular, clone 8 has
            been inoculated intravenously into two pig-tailed macaques causing
            CD4 lymphocyte depletion; see J. Virol. 62, 2091-2101, 1988.  The
            env cds is truncated as is the case with MM251 and MM142.
               Author address:R.Benveniste
               National Cancer Institute
               Frederick, MD
               (301-698-5836).
FEATURES             Location/Qualifiers
     source          1. .9628
                     /organism="Simian immunodeficiency virus"
                     /proviral
                     /specific_host="Macaca nemestrina"
                     /db_xref="taxon:11723"
                     /clone="8."
     CDS             533. .2053
                     /note="gag polyprotein"
                     /codon_start=1
                     /protein_id="AAA91930.1"
                     /db_xref="GI:334696"
                     /translation="MGARNSVLSGKKADELEKIRLRPGGKKKYMLKHVVWAANELDRF
                     GLAESLLENKEGCQKILSVLAPLVPTGSENLKSLYNTVCVIWCIHAEEKVKHTEEAKQ
                     IVQRHLVVETGTAETMPKTSRPTAPSSGRGGNYPVQQVGGNYTHLPLSPRTLNAWVKL
                     IEEKKFGAEVVPGFQALSEGCTPYDINQMLNCVGEHQAAMQIIREIINEEAADWDLQH
                     PQQAPQQGQLREPSGSDIAGTTSTVDEQIQWMYRQQNPIPVGNIYRRWIQLGLQKCVR
                     MYNPTSILDVKQGPKEPFQSYVDRFYKSLRAEQTDPAVKNWMTQTLLIQNANPDCKLV
                     LKGLGMNPTLEEMLTACQGVGGPGQKARLMAEALKEALAPGPLPFAAAQQKGPRKPIK
                     CWNCGKEGHSARQCRTPRRQGCWKCGQMGHVMAKCPDRQAGFLGFGPWGKKPRNFPMA
                     QMHQGLTPTAPPEDPAVDLLKNYMQLGKQQRESKRKPYKEVTEDLLHLNSLFGEDQ"
     CDS             1708. .4878
                     /partial
                     /note="NH2-terminus uncertain"
                     /codon_start=1
                     /product="pol polyprotein"
                     /protein_id="AAA91931.1"
                     /db_xref="GI:334697"
```

```
                     /translation="VLELWERGTLCKAMQNPKKTGLLEMWTNGPCYGQMPRQTGRFFR
                     LWPMGKEAPQFPHGPNASGADANCSPRGPSCGSAKKLHAVGQTAERKQKEALQGGDRG
                     FAAPQFSLWRRPVVTAHIEGQPAEVLLDTGADDSIVAGIELGPHYTPKIVGGIGGFIN
                     TKEYKNVKIEVLGKRIKGTIMTGDTPINIFGRNLLTALEMSLNFPIAKVEPVKVTLKP
                     GKDGPKLRQWPLSKEKIVALREICEKMEKDGQLEEAPPTNPYNTPTFAIKKKDKNKWR
                     ILIDFRELNKVTQDFTEVQLGIPHPAGLAKRKRITVLDVGDAYFSIPLDEEFRQYTAF
                     TLPSVNNAEPGKRYIYKVLPQGWKGSPAIFQHTMRNVLEPFRKANPDVTLVQYMDDIL
                     VASDRTDLEHDRVVLQLKELLNSIGFSTPEEKFQKDPPFQWMGYELWPTKWKLQKIEL
                     PQKETWTVNDIQKLVGVLNWAAQIYPGIKTKHLCRLIRGKMTLTEEVQWTEMAEAEYE
                     ENKIILSQEQEGCYYQEGKPLEATVIKNQDNQWSYKIHQEDKILKVGKFAKIKNTHTN
                     GVRLLAHVIQKIGKEAIVIWGQVPKFHLPVEKDVWEQWWTDYWQVTWIPKWDFISTPP
                     LVRLVFNLVKDPIKGEETYYVDGSCNKQSKEGKAGYITDRGKDKVKVLEQTTNQQAEL
                     EAFLMALADSGPKANIIVDSQYVMGIITGCPTESESRLVNQIIEEMIKKTEIYVAWVP
                     AHKGIGGNQEIDHLVSQGIRQVLFLEKIEPAQEEHDKYHSNVKELVFKFGLPRLVAKQ
                     IVDTCDKCHQKGEAIHGQVNSDLGTWQMDCTHLEGKIIIVAVHVASGFIEAEVIPQET
                     GRQTALFLLKLASRWPITHLHTDNGANFASQEVKMVAWWAGIEHTFGVPYNPQSQGVV
                     EAMNHHLKNQIDRIREQANSMETIVLMAVHCMNFKRRGGIGDMTPAERLLNMITTEQE
                     IQFQQSKNSKFKNFRVYYREGRDQLWKGPGELLWKGEGAVVLKVGTDIKVVPRRKAKI
                     IKDYGGGKEVDSSSHMEDTGEAREVA"
CDS                  4808. .5452
                     /note="vif protein"
                     /codon_start=1
                     /protein_id="AAA91932.1"
                     /db_xref="GI:334698"
                     /translation="MEEEKRWIAVPTWRIPERLERWHSLIKYLKYKTKDLQKVCYVPH
                     HKVGWAWWTCSRVIFPLQEKSQLEVQGYWNLTPERGWLSTHAVRITWYSRNFWTDVTP
                     DCADILLHSTYFPCFTAGEVRRAIRGEQLLSCCRFPRAHKTQVPSLQYLALRVVSYVR
                     SQRENPTWKQWRRDNRRSLRMAKQNSRGDKQRGSKPPTKGVDFPGLAKVLGILA"
CDS                  5280. .5618
                     /note="vpX protein"
                     /codon_start=1
                     /protein_id="AAA91933.1"
                     /db_xref="GI:334699"
                     /translation="MSDPRERIPPGNSGEETIGEAFEWLNRTVEEINREAVNHLPREL
                     IFQVWQRSWEYWHDEQGMSQSYVKYRYLCLIQKALFMHCKKGCRCLGEGHGAGGWRPG
                     PPPPPPPGLA"
CDS                  5619. .5924
                     /note="vpR protein"
                     /codon_start=1
                     /protein_id="AAA91934.1"
                     /db_xref="GI:334700"
                     /translation="MEERPPEDEGPQREPWDEWVVEVLEELKEEALKHFDPRLLTALG
                     NHIYNRHGDTLEGAGELIKILQRALFMHFRGGCTHSRIGQSGGGNPLSTIPPSRRML"
CDS                  join(5770. .6065,8280. .8376)
                     /note="tat protein"
                     /codon_start=1
                     /protein_id="AAA91935.1"
                     /db_xref="GI:334693"
                     /translation="METPLREQENSLKSSNGRSSCTSEAAAPTLESANLEEEILSQLY
                     RPLEECYNTCYCKKCCYHCQFCFLKKGLGICYEQSRRRRRTPKKAKANTSSASNKPIS
                     DRTRHCQPKKEQKETVEAAVATAPGLGR"
CDS                  join(5996. .6065,8280. .8533)
                     /note="rev protein"
                     /codon_start=1
                     /protein_id="AAA91936.1"
                     /db_xref="GI:334694"
                     /translation="MSSHAEEEELRRRLRLIHLLHQTNPYPTGPGTANQRRNRRRRWR
                     QRWQQLLALADRIYSFPDPPTNTPLDLVIQQLQNLAIESIPDPPTNIPEILHDPTENP
                     RSPQD"
CDS                  6072. .8777
                     /note="env protein (premature stop codon)"
                     /pseudo
                     /codon_start=1
CDS                  join(6072. .8276,8280. .8717)
                     /note="env protein"
                     /codon_start=1
                     /protein_id="AAA91937.1"
                     /db_xref="GI:334695"
                     /translation="MGCLGNQLLIAILFLSAYGIYCIQYVTVFYGVPAWRNATIPLFC
                     VTRNRDTWGTTQCLPDNDDYSELALNITESFDAWENTVTEQAIEDVWHLFETSIKPCV
                     KLTPLCITMKCNKSETDKWGLTKSSTTTAPTAIPTKAEAIKVVNENSPCINHDNCTGL
                     EQEPMISCKFNMTGLKRDKRREYNETWYSADLVCEQGNSTENESRCYMNHCNTSVIQE
                     SCDKHYWDAIRFRYCAPPGYALLRCNDTNYSGFMPNCSKVVVSSCTRMMETQTSTWFG
                     FNGTRAENRTYIYWHSKDNRTIISLNKYYNLTMKCRRPGNKTVLPVTIMSGLVFHSQP
                     INDRPKQAWCRFEGNWKEAIKEVKQTIVKHPRYTGTNNTDKINLTAPGGGDPEVTFMW
                     TNCRGEFLYCKMNWFLNWVEDKNLTGTTQKPQERHKRNYVPCHIRQIINTWHKVGRNV
                     YLPPREGDLTCNSTVTSLIANIDWIDGNQTNITMSAEVAELYRLELGDYKLVEITPIG
                     LAPTNVKRYTTGGTSRNKRGVFVLGFLGFLATAGSAMGAASLTLTAQSRTLLAGIVQQ
                     QQQLLDVVKRQQELLRLTVWGTKNLQTRVTAIEKYLKDQAQLNAWGCAFRQVCHTTVP
                     WPNANLTPNWNNETWQEWERKVDFLEENITALLEEAQIQQEKNMYELQKLNSWDVFGN
                     WFDLASWIRYIQYGVYIVVGVILLRIVIYIVQMLAKLRQGYRPVFSSPPSYFQTHIRQ
                     DQALPTKEGTEGDGGGSGGNSSWPWQIEYIHFLIRQLIRLLTWLFSNCRTLLSRAYQI
                     LQPIFQRFSTTLQRIREVLRTELTYLQYGWSYFQEAVQVAWRSATETLAGAWGDLWET
                     LGRVGRWILAIPRRIRQGLELTLL"
misc_feature         8277. .8279
                     /note="env protein in-frame stop codon"
CDS                  8551. .9342
                     /note="nef protein"
                     /codon_start=1
                     /protein_id="AAA91938.1"
                     /db_xref="GI:334701"
                     /translation="MGGATSKRRSKSPGDLRQRLLRARGETYGRLWEELEDGYSQSLG
                     GSDKGSSLLSCEGQKYNQGQFMNTPWKNPAGEREKLAYRKQNIDDIEEDNDLVGVPV
                     RPRVPLRIISYKLAVDMSHFIKEKGGLEGIYYSERRHKILDMYLEKEEGIMPDWQNYT
                     SGPGPRYPKTFGWLWKLVPVNVSDEAQEGEENYLLHPAQTSQWDDPWGEVLVWKFDPT
                     LAYTYEAYIRYPEEFGSKSGLSEEEVRRRLTARGLLKMADKRETS"
BASE COUNT     3294 a    1807 c    2379 g    2148 t
ORIGIN      5' end of 5' LTR R region (putative mRNA start).

  Query Match             92.7%;  Score 1772.6;  DB 14;  Length 9628;
  Best Local Similarity   96.2%;  Pred. No. 0;
  Matches 1841;  Conservative    0;  Mismatches   59;  Indels   13;  Gaps    2;

Qy       1 GGTACCAGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGCAAC 60
             ||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db     418 GGTACCAGACGGCGTGAGGAGCGGGAGA-GAAGAGGCCTCCGGTTGCAGGTAAGTGCAAC 476

Qy      61 ACAAAAAAGAAATAGCTGTCTTTTATCCAGGAAGGGGTAATAAGATAGAGTGGGAGATGG 120
             |||||||||| |||||||||||||||||||| |||||||||||||||||||||||||||
Db     477 ACAAAAAAGAGATAGCTGTCTTTTATCCAGGAAGGGATAATAAGATAGAGTGGGAGATGG 536

Qy     121 GCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCTAC 180
             ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     537 GCGCGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCTAC 596

Qy     181 GACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTAG 240
             |||||  ||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     597 GACCCGGCGGGAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTAG 656

Qy     241 ATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTTT 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     657 ATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTTT 716

Qy     301 CGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACTG 360
             |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     717 CGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAGAGCCTTTATAATACTG 776

Qy     361 TCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAAC 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     777 TCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAAC 836

Qy     421 AGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAAACAA 480
             |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     837 AGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAGCAGAAACTATGCCAAAAACAA 896

Qy     481 GTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGGTG 540
             |||||||||||||||||||| |||||||||||||||||||||||||| ||||||||||||
Db     897 GTAGACCAACAGCACCATCTAGTGGCAGAGGAGGAAATTACCCAGTACAACAAGTAGGTG 956

Qy     541 GTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGATAG 600
             |||||||| |||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     957 GTAACTATACCCACCTACCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGATAG 1016

Qy     601 AGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGTTGCA 660
             |||| ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db    1017 AGGAGAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGCTGCA 1076
```

```
Qy     661 CCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCAGA 720
           ||||||||||||||||||||||||||||||||||||||||||| |||||||| ||||||||||
Db    1077 CCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGAACATCAAGCAGCTATGCAGA 1136

Qy     721 TTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACCAG 780
           |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db    1137 TTATCAGAGAAATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACAAG 1196

Qy     781 CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACTAGTT 840
           ||||||||||||||| |||||||||||||||||||||| ||||||||||||||||||||||
Db    1197 CTCCACAACAAGGACAGCTTAGGGAGCCGTCAGGATCAGACATTGCAGGAACAACTAGTA 1256

Qy     841 CAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGCAACA 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1257 CAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGCAACA 1316

Qy     901 TTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCAACAA 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1317 TTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCAACAA 1376

Qy     961 ACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCT 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1377 GCATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCT 1436

Qy    1021 ACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACAC 1080
           ||||||| ||||||||||||||| |||||| ||||||||||||||||||||||||||||||
Db    1437 ACAAAAGCTTAAGAGCAGAACAAACAGATCCAGCAGTAAAGAATTGGATGACTCAAACAC 1496

Qy    1081 TGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATC 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db    1497 TGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTATGAATC 1556

Qy    1141 CCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCTA 1200
           |||||||||||||||||||||||||||||||||||||||||| || || ||||| ||||| |
Db    1557 CCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGAGGACCAGGACAAAAGGCAA 1616

Qy    1201 GATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCAGCAG 1260
           |||||||||||||||||||||||||||||| ||||||| |||| ||||||||||||||||
Db    1617 GATTAATGGCAGAAGCCCTGAAAGAGGCCCTTGCACCAGGGCCACTCCCTTTTGCAGCAG 1676

Qy    1261 CCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGACACT 1320
           ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1677 CCCAACAGAAGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGACACT 1736

Qy    1321 CTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATGGACC 1380
           |||||||||||||||| |||||||||||||||| |||||||||||||| |||||| ||||
Db    1737 CTGCAAGGCAATGCAGAACCCCAAGAAGACAGGGCTGCTGGAAATGTGGACAAATGGGCC 1796

Qy    1381 ATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGGGGAA 1440
           |||||||||||||||||||||||||||||||||||||||||||||||||| |||| ||||||||||
Db    1797 ATGTTATGGCCAAATGCCCAGACAGACAGGCAGGTTTTTTAGGCTTTGGCCCATGGGGAA 1856

Qy    1441 AGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGCTCCCC 1500
           |||||||||||||||||||||||| ||| |||||||||||||| ||||||||||||||||||
Db    1857 AGAAGCCCCGCAATTTCCCCATGGCCCAAATGCATCAGGGGCTGACGCCAACTGCTCCCC 1916

Qy    1501 CAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAGAGAG 1560
           |||||||||||||||||||||||||||||| ||||||||||||||||||
Db    1917 CAGAGGACCCAGCTGTGGATCTGCTAAAAAACTACATGCAGTTGGGCA------------ 1964

Qy    1561 AAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCACC 1620
           || |||||||||||||| |  ||||||||||||||||||||||||||||||||||||||||
Db    1965 AACAGCAGAGAGAAAGCAAAAGGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCACC 2024

Qy    1621 TCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAGA 1680
           |||||||||||||||| |||||||||||||||||||||||||| |||||||||| |||||
Db    2025 TCAATTCTCTCTTTGGAGAAGACCAGTAGTCACTGCTCATATTGAGGGACAGCCTGCAGA 2084

Qy    1681 AGTATTACTGGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCACA 1740
           ||||||| | ||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    2085 AGTATTATTAGATACAGGGGCTGATGATTCTATTGTAGCAGGAATAGAGTTAGGTCCACA 2144

Qy    1741 TTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAAA 1800
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2145 TTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAAA 1800

Qy    1801 TGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCCC 1860
           |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2205 TGTAAAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCCC 1860

Qy    1861 GATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAATT 1913
           |||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    2265 GATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTGGAGATGTCTCTAAATT 2317
```

```
RESULT 14
SIU79412
LOCUS       SIU79412                 9664 bp    DNA     linear   VRL 21-JAN-1998
DEFINITION  Simian immunodeficiency virus isolate SIVMne027, complete genome.
ACCESSION   U79412
VERSION     U79412.1  GI:2737927
KEYWORDS    .
SOURCE      Simian immunodeficiency virus.
  ORGANISM  Simian immunodeficiency virus
            Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
            lentivirus group.
REFERENCE   1  (bases 1 to 9664)
  AUTHORS   Kimata,J.T., Mozaffarian,A. and Overbaugh,J.
  TITLE     A lymph node-derived cytopathic simian immunodeficiency virus Mne
            variant replicates in nonstimulated peripheral blood mononuclear
            cells
  JOURNAL   J. Virol. 72 (1), 245-256 (1998)
  MEDLINE   98080410
  PUBMED    9420221
REFERENCE   2  (bases 1 to 9664)
  AUTHORS   Kimata,J.T., Mozaffarian,A. and Overbaugh,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-1996) Microbiology, University of Washington,
            Health Sciences Building, HSB 320, Seattle, WA 98195-7242, USA
FEATURES             Location/Qualifiers
     source          1. .9664
                     /organism="Simian immunodeficiency virus"
                     /proviral
                     /strain="Mne"
                     /isolate="SIVMne027"
                     /db_xref="taxon:11723"
     LTR             <1. .300
                     /note="5' long terminal repeat"
     repeat_region   1. .175
                     /note="R region"
                     /rpt_type=direct
     repeat_region   176. .300
                     /note="U5 region"
                     /rpt_type=direct
     gene            533. .2053
     CDS             /gene="gag"
                     533. .2053
                     /gene="gag"
                     /codon_start=1
                     /product="gag polyprotein"
                     /protein_id="AAB97129.1"
                     /db_xref="GI:2737928"
                     /translation="MGARNSVLSGKKADELEKIRLRPGGKKKYMLKHVVWAANELDRF
                     GLADSLLENKEGCQKILSALAPLVPTGSENLKSLYNTVCVIWCIHAEEKVKHTEEAKQ
                     IVQRHLVVETGTAETMPKTSRPTAPSSGKRGNYPVQQIGGNYTHLPLSPRTLNAWVKL
                     IEEKKFGAEVVPGFQALSEGCIPYDINQMLNCVGEHQAAMQIIREIINEEAADWDLQH
                     PQQAPQQGQLREPSGSDIAGTTSTVDEQIQWMYRQQNPIPVGNIYRRWIQLGLQKCVK
                     MYNPTNILDIRQGPKEPFQSYVDRFYKSLRAEQTDPAVKNWMTQTLLIQNANPDCKLV
                     LKGLGMNPTLEEMLTACQGVGGPGQKAKLMAEALKEALAPGPLPFAAAQQKGPRKPIK
                     CWNCGKEGHSARQCRAPRRQGCWKCGQMGHVMAKCPDRQAGFLGFGPWGKKPRNFPMA
                     QMHQGLTPTAPPEDPAVDLLKNYMQLGKQQRESKRKPYKEVAEDLLHLNSLFGEDQ"
     gene            1918. .4878
     CDS             /gene="pol"
                     <1918. .4878
                     /gene="pol"
```

```
                     /codon_start=1
                     /product="pol polyprotein"
                     /protein_id="AAB97130.1"
                     /db_xref="GI:2737929"
                     /translation="RGPSCGSAKKLHAVGQTAERKQKEALQGGGRGFAAPQFSLWRRP
                     VVTAHIEGQPAEVLLDTGADDSIVAGIELGPHYTPKIVGGIGGFINTKEYKNVEIEVL
                     GKRIKGTIMTGDTPINIFGRNLLTALGMSLNFPIAKVEPVKVTLKPGKDGPKLRQWPL
                     SKEKIVALREICEKMEKDGQLEEAPPTNPYNTPTFAIKKKDKNKWRMLIDFRELNKVT
                     QDFTEVQLGIPHPAGLAKRKRITVLDVGDAYFSIPLDEEFRQYTAFTLPSVNNAEPGK
                     RYIYKVLPQGWKGSPAIFQHTMRNVLEPFRKANPDVTLVQYMDDILVASDRTDLEHDR
                     VVLQLKELLNSIGFSTPEEKFQKDPPFQWMGYELWPTKWKLQKIELPQKETWTVNDIQ
                     KLVGVLNWAAQIYPGIKTKHLCRLIRGKMTLTEEVQWTEMAEAEYEENKIILSQEQEG
                     CYYQEGKPLEATVIKNQDNQWSYKIHQEDKILKVGKFAKIKNTHTNGVRLLAHVIQKI
                     GKEAIVIWGQVPKFHLPVEKDVWEQWWTDYWQVTWIPEWDFISTPPLVRLVFNLVKDP
                     IEGEETYYVDGSCNKQSKEGKAGYITDRGKDKVRVLEQTTNQQAELEAFLMALTDSGP
                     KANIIVDSQYVMGIITGCPTESESRLVNQIIEEMIKKTEIYVAWVPAHKGIGGNQEID
                     HLVSQGIRQVLFLEKIEPAQEEHDKYHSNVKELVFKFGLPRLVAKQIVDTCDKCHQKG
                     EAIHGQVNSDLGTWQMDCTHLEGKIIIVAVHVASGFIEAEVIPQETGRQTALFLLKLA
                     SRWPITHLHTDNGANFTSQEVKMVAWWAGIEHTFGVPYNPQSQGVVEAMNHHLKNQID
                     RIREQANSMETIVLMAVHCMNFKRRGGIGDMTPAERLLNMITTEQEIQFQQSKNSKFK
                     NFRVYYREGRDQLWKGPGELLWKGEGAVILKVGTDIKVVPRRKAKIIKDYGGGKEVDS
                     SSHMEDTGEAREVA"
gene                 4808. .5452
                     /gene="vif"
CDS                  4808. .5452
                     /gene="vif"
                     /codon_start=1
                     /product="vif protein"
                     /protein_id="AAB97131.1"
                     /db_xref="GI:2737930"
                     /translation="MEEEKRWIAVPTWRIPERLERWHSLIKYLKYKTKDLQKVCYVPH
                     HKVGWAWWTCSRVIFPLQEESQLEVQGYWNLTPERGWLSTYAVRITWYSRNFWTDVTP
                     DYADILLHSTYFPCFTAGEVRRAIRGEQLLSCCRFPRAHKNQVPSLQYLALRVVSYVR
                     SQRENPTWKQWRRDNRRSLRMAKQNSRGDKQRGSKPPTKGADFPGLAKVLGILA"
gene                 5280. .5618
                     /gene="vpx"
CDS                  5280. .5618
                     /gene="vpx"
                     /codon_start=1
                     /product="vpx protein"
                     /protein_id="AAB97132.1"
                     /db_xref="GI:2737931"
                     /translation="MSDPRERIPPGNSGEETIGEAFEWLNRTVEEINREAVNHLPREL
                     IFQVWQRSWEYWHDEQGMSPSYVKYRYLCLIQKALFMHCKKGCRCLGEGHGAGGWRPG
                     PPPPPPPGLA"
gene                 5619. .5924
                     /gene="vpr"
CDS                  5619. .5924
                     /gene="vpr"
                     /codon_start=1
                     /product="vpr protein"
                     /protein_id="AAB97133.1"
                     /db_xref="GI:2737932"
                     /translation="MEERPPEDEGPQREPWDEWVVEVLEELKEEALKHFDPRLLTALG
                     NYIYDRHGDTLEGAGELIKILQRALFMHFRGGCNHSRIGQSGGGNPLSTIPPSRSML"
gene                 5770. .8379
                     /gene="tat"
CDS                  join(5770. .6065,8283. .8379)
                     /gene="tat"
                     /codon_start=1
                     /product="tat protein"
                     /protein_id="AAB97134.1"
                     /db_xref="GI:2737933"
                     /translation="METPLREQENSLKSSNGRSSCTSEAAATTLESANLEEEILSQLY
                     RPLEACYNTCYCKKCCYHCQFCFLKKGLGICYEQSRRRRRTPKKVKANTSSASNKPIS
                     DRTKHCQPKKEQKETVETAVATVPGLGR"
gene                 5996. .8536
                     /gene="rev"
CDS                  join(5996. .6065,8283. .8536)
                     /gene="rev"
                     /codon_start=1
                     /product="rev protein"
                     /protein_id="AAB97135.1"
                     /db_xref="GI:2737934"
                     /translation="MSSHAEEEELRRRLRLIHLLHQTNPYPTGPSTANQRRNRRRRWR
                     QRWQQFLALADRVYSFPDPPTDTPLDLAIQQLQNLAIESIPDPPTNIPEILHDPTESP
                     RSPQD"
gene                 6072. .8720
                     /gene="env"
CDS                  6072. .8720
                     /gene="env"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAB97136.1"
                     /db_xref="GI:2737935"
                     /translation="MGCLGNQLLIAILFLSAYGIYCIQYVTVFYGVPAWRNATIPLFC
                     VTRNRDTWGTTQCLPDNDDYSELALNITESFDAWENTVTEQAIEDVWHLFETSIKPCV
                     KLTPLCITMKCNKSETDKWGLTKSSTTTAAPTTKTTTTKEIEVVNENSTCVNRDNCTG
                     LEQEPMISCKFNMTGLKRDKKREYNETWYSADLVCEQGNSTEDESRCYMNHCNTSVIQ
                     ESCDKHYWDAIRFRYCAPPGYALLRCNDTKYSGFMPNCSKVVVSSCTRMMETQTSTWF
                     GFNGTRAENRTYIYWHSKDNRTIISLNKYNNLTMKCRRPGNKTVLPVTIMSGLVFHSQ
                     PINERPKQAWCRFEGNWKEAIKEVKQTIVKHPRYTGTNNTDKINLTAPRGGDPEVTFM
                     WTNCRGEFLYCKMNWFLNWVEDKNLTGTTQKPQEQHKRNYVPCHIRQIINTWHKVGKN
                     VYLPPREGDLTCNSTVTSLIANIDWIDGNQTNITMSAEVAELYRLELGDYKLVEITPI
                     GLAPTNVKRYTTGGTPRNKRGVFVLGFLGFLATAGSAMGAASLTLTAQSRTLLAGIVQ
                     QQQQLLDVVKRQQELLRLTVWGTKNLQTRVTAIEKYLKDQAQLNAWGCAFRQVCHTTV
                     PWPNASLTPNWNNETWQEWERKVDFLEENITALLEEAQIQQEKNMYELQKLNSWDVFG
                     NWFDLASWIRYIQYGVYIVVGVILLRIVIYIVQMLAKLRQGYRPVFSSPPSYFQQTHI
                     RQDQALPTKEGTEGDGGDSGGNSSWPWQIEYIHFLIRQLIRLLTWLFSNCRTLLSRAY
                     QILQPIFQRFSTTLQRVREVLRTELTYLQYGWSYFQEAVQVAWRSATETLAGAWGDLW
                     ETLGRVGRWILAIPRRIRQELELTLL"
gene                 8554. .9345
                     /gene="nef"
CDS                  8554. .9345
                     /gene="nef"
                     /codon_start=1
                     /product="nef protein"
                     /protein_id="AAB97137.1"
                     /db_xref="GI:2737936"
                     /translation="MGGATSKKRSKSPGDLRQRLLRARGETYGRLWEGLEDGYSQSLG
                     GSDKSLSLLSCEGQKYNQGQFMNTPWKNPAGEREKLAYRKQNMDDIEEDDGLVGVPV
                     RPRVPLRAMTYKLAVDMSHFIKEKGGLEGIYYSERRHKILDIYLEKEEGIVPDWQDYT
                     SGPGPRYPKTFGWLWKLVPVNVSDEAQEGEENYLLHPAQTSQWDDPWGEVLVWKFDPT
                     LAYTYEAYIRYPEEFGSKSGLSEEEVRRRLTARGLLKMADKKETS"
LTR                  8939. .>9664
                     /note="3' long terminal repeat"
repeat_region        8939. .9457
                     /note="U3 region"
                     /rpt_type=direct
repeat_region        9457. .9632
                     /note="R region"
                     /rpt_type=direct
repeat_region        9633. .9664
                     /note="U5 region"
                     /rpt_type=direct
BASE COUNT     3301 a   1829 c   2387 g   2147 t
ORIGIN

  Query Match             91.8%;  Score 1756.6;  DB 14;  Length 9664;
  Best Local Similarity   95.7%;  Pred. No. 0;
  Matches 1831;  Conservative    0;  Mismatches   69;  Indels   13;  Gaps    2;

Qy       1 GGTACCAGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGCAAC 60
           |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db     418 GGTACCAGACGGCGTGAGGAGCGGGAGA-GAAGAGGCCTCCGGTTGCAGGTAAGTGCAAC 476

Qy      61 ACAAAAAAGAAATAGCTGTCTTTTATCCAGGAAGGGGTAATAAGATAGAGTGGGAGATGG 120
           ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db     477 ACAAAAAAGAAATAGCTGTCTTTTATCCAGGAAGGGATAATAAGATAGAGTGGGAGATGG 536

Qy     121 GCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCTAC 180
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     537 GCGCGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCTAC 596

Qy     181 GACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTAG 240
           |||||  ||| ||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    597 GACCCGGCGGGAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTAG 656

Qy    241 ATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTTT 300
          ||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db    657 ATAGATTTGGATTAGCAGACAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTTT 716

Qy    301 CGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACTG 360
          ||| ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db    717 CGGCCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAGAGCCTTTATAATACTG 776

Qy    361 TCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAAC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    777 TCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAAC 836

Qy    421 AGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAAACAA 480
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    837 AGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAGCAGAAACTATGCCAAAAACAA 896

Qy    481 GTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGGTG 540
          |||||||||||||||||||| |||| | ||||||||||||||||||||||||||||||||
Db    897 GTAGACCAACAGCACCATCTAGTGGCAAAAGAGGAAATTACCCAGTACAACAAATAGGTG 956

Qy    541 GTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGATAG 600
          |||||||| |||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    957 GTAACTATACCCACCTACCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGATAG 1016

Qy    601 AGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGTTGCA 660
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db   1017 AGGAGAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGCTGCA 1076

Qy    661 CCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCAGA 720
          |||||||||||||||||||||||||||||||||||||| |||||||| ||||||||||||
Db   1077 TCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGAACATCAAGCAGCTATGCAGA 1136

Qy    721 TTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACCAG 780
          |||||||||| |||||||||||||||||||||||||||||||| ||||||||||||||| ||
Db   1137 TTATCAGAGAAATTATAAACGAGGAGGCTGCAGATTGGGACCTGCAGCACCCACAACAAG 1196

Qy    781 CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACTAGTT 840
          |||||||||||||||| |||||||||||||||||||||| ||||||||||||||||||||
Db   1197 CTCCACAACAAGGACAGCTTAGGGAGCCGTCAGGATCAGACATTGCAGGAACAACTAGTA 1256

Qy    841 CAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGCAACA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1257 CAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGCAACA 1316

Qy    901 TTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCAACAA 960
          |||| |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   1317 TTTATAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAAAATGTATAACCCAACAA 1376

Qy    961 ACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCT 1020
          |||||||||| ||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   1377 ACATTCTAGATATAAGACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCT 1436

Qy   1021 ACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACAC 1080
          ||||||| |||||||||||||| |||||| ||||||||||||||||||||||||||||||
Db   1437 ACAAAAGCTTAAGAGCAGAACAAACAGATCCAGCAGTAAAGAATTGGATGACTCAAACAC 1496

Qy   1081 TGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATC 1140
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||| ||||||
Db   1497 TGCTGATTCAAAATGCTAACCCAGATTGCAAGCTGGTGCTGAAGGGGCTGGGTATGAATC 1556

Qy   1141 CCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCTA 1200
          |||||||||||||||||||||||||||| |||||||||| || || ||||| |||||| |
Db   1557 CCACCCTAGAAGAAATGCTGACGGCTTGCCAAGGAGTAGGAGGACCAGGACAAAAGGCAA 1616

Qy   1201 GATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCAGCAG 1260
          |||||||||||||||||||||||||||||||| ||||||| ||| ||||||||||||||||
Db   1617 AATTAATGGCAGAAGCCCTGAAAGAGGCCCTTGCACCAGGGCCACTCCCTTTTGCAGCAG 1676

Qy   1261 CCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGACACT 1320
          ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1677 CCCAACAGAAGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGACACT 1736

Qy   1321 CTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATGGACC 1380
          |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||| ||
Db   1737 CTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGCTGCTGGAAATGTGGACAAATGGGCC 1796

Qy   1381 ATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGGGGAA 1440
          |||||||||||||||||||||||||||||||| |||||||||||| |||| ||||||||||
Db   1797 ATGTTATGGCCAAATGCCCAGACAGACAGGCAGGTTTTTTAGGCTTTGGCCCATGGGGAA 1856

Qy   1441 AGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGCTCCCC 1500
          |||||||||||||||||||||||| ||| |||||||||||||| ||||||||||||||||
Db   1857 AGAAGCCCCGCAATTTCCCCATGGCCCAAATGCATCAGGGGCTGACACCAACTGCTCCCC 1916

Qy   1501 CAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAGAGAG 1560
          |||||||||||||||||||||||||||||| ||||||||||||||||||
Db   1917 CAGAGGACCCAGCTGTGGATCTGCTAAAAAACTACATGCAGTTGGGCA---------- 1964

Qy   1561 AAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCACC 1620
          || |||||||||||||| | |||||||||||||||||||| ||||||||||||||||||||
Db   1965 AACAGCAGAGAGAAAGCAAAAGGAAGCCTTACAAGGAGGTGGCAGAGGATTTGCTGCACC 2024

Qy   1621 TCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAGA 1680
          |||||||||||||||||| ||||||||||||||||||||||||| |||||||||| |||
Db   2025 TCAATTCTCTCTTTGGAGAAGACCAGTAGTCACTGCTCATATTGAGGGACAGCCTGCAGA 2084

Qy   1681 AGTATTACTGGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCACA 1740
          |||| || ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db   2085 AGTACTATTGGATACAGGGGCTGATGATTCTATTGTAGCAGGAATAGAGTTAGGTCCACA 2144

Qy   1741 TTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAAA 1800
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2145 TTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAAA 2204

Qy   1801 TGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCCC 1860
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2205 TGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCCC 2264

Qy   1861 GATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAATT 1913
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   2265 GATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTGGGAATGTCTCTAAATT 2317
```

```
RESULT 15
SIVJ11P55
LOCUS       SIVJ11P55                1533 bp    DNA     linear   VRL 09-MAY-1996
DEFINITION  Simian immunodeficiency virus (32H Re-isolate of SIVmac 251, clone
            J11) gag gene, complete cds.
ACCESSION   M74934
VERSION     M74934.1  GI:334584
KEYWORDS    retroviral structural protein.
SOURCE      Simian immunodeficiency virus (tissue library: 11/88 stock) DNA.
  ORGANISM  Simian immunodeficiency virus
            Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
            lentivirus group.
REFERENCE   1  (bases 1 to 1533)
  AUTHORS   Almond,N., Jenkins,A., Slade,A., Heath,A., Cranage,M. and
            Kitchin,P.
  TITLE     Population sequence analysis of a simian immunodeficiency virus
            (32H reisolate of SIVmac251): a virus stock used for international
            vaccine studies
  JOURNAL   AIDS Res. Hum. Retroviruses 8 (1), 77-88 (1992)
  MEDLINE   92144212
   PUBMED   1736942
FEATURES             Location/Qualifiers
     source          1. .1533
                     /organism="Simian immunodeficiency virus"
                     /proviral
                     /db_xref="taxon:11723"
                     /cell_line="C8166"
                     /tissue_lib="11/88 stock"
     gene            complement(1. .1533)
                     /gene="gag"
     CDS             1. .1533
```

```
                     /codon_start=1
                     /product="p55"
                     /protein_id="AAA99258.1"
                     /db_xref="GI:334585"
                     /translation="MGARNSVLSGKKADELEKIRLRPGGKKKYMLKHVVWAANELDRF
                     GLAESLLENKEGCQKILSVLAPLVPTGSENLKSLYNTVCVIWCIHAEEKVKHTEEAKQ
                     IVQRHLVVETGTAETMPKTSRPTAPSSGRGGNYPVQQIGGNYVHLPLSPRTLNAWVKL
                     IEEKKFGAEVVPGFQALSEGCTPYDINQMLNCVGDHQAAMQIIRDIINEEAADWDLQH
                     PQPAPQQGQLREPSGSDIAGTTSSVDEQIQWMYRQQNPIPVGNIYRRWIQLGLQKCVR
                     MYNPTNILDVKQGPKEPFQSYVDRFYKSLRAEQTDAAVKNWMTQTLLIQNANPDCKLV
                     LKGLGVNPTLEEMLTACQGVGGPGQKARLMAEALKEALAPVPIPFAAAQQRGPRKPIK
                     CWNCGKEGHSARQCRAPRRQGCWKCGKMDHVMAKCPDRQAGFLGLGPWGKKPRNFPMA
                     QVHQGLTPTAPPEDPAVDLLKNYMQLGKQQREKQRESREKPYKEVTEDLLHLNSLFGG
                     DQ"
     primer_bind     complement(1. .21)
                     /partial
                     /gene="gag"
                     /note="5' PCR primer"
     primer_bind     complement(1513. .1533)
                     /partial
                     /gene="gag"
                     /note="3' PCR primer"
BASE COUNT      532 a    314 c    401 g    286 t
ORIGIN
```

Query Match 78.8%; Score 1507.4; DB 14; Length 1533;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1517; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

```
Qy    117 ATGGGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGG 176
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGGGCGCGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGG 60

Qy    177 CTACGACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAA 236
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CTACGACCCGGCGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAA 120

Qy    237 TTAGATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATA 296
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TTAGATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATA 180

Qy    297 CTTTCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAAT 356
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CTTTCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAAT 240

Qy    357 ACTGTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCA 416
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ACTGTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCA 300

Qy    417 AAACAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAA 476
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    301 AAACAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAGCAGAAACTATGCCAAAA 360

Qy    477 ACAAGTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATA 536
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ACAAGTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATA 420

Qy    537 GGTGGTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTG 596
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GGTGGTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTA 480

Qy    597 ATAGAGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGT 656
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ATAGAGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGC 540

Qy    657 TGCACCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATG 716
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TGCACCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATG 600

Qy    717 CAGATTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAA 776
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    601 CAGATTATCAGAGACATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAA 660

Qy    777 CCAGCTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACT 836
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    661 CCAGCTCCACAACAAGGACAGCTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACT 720

Qy    837 AGTTCAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGC 896
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 AGCTCAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGC 780

Qy    897 AACATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCA 956
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 AACATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCA 840

Qy    957 ACAAACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGG 1016
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    841 ACAAACATTCTAGATGTAAAACAAGGGCCAAAGGAGCCATTTCAGAGCTATGTAGACAGG 900

Qy   1017 TTCTACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAA 1076
          |||||||||| |||||||||||||| ||||||||||||||||||||||||||||||||||
Db    901 TTCTACAAAAGCTTAAGAGCAGAACAAACAGATGCAGCAGTAAAGAATTGGATGACTCAA 960

Qy   1077 ACACTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTG 1136
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 ACACTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTG 1020

Qy   1137 AATCCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAG 1196
          |||||||||||||||||||||||||||||||||||||||||||||| || ||||||||||
Db    1021 AATCCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGACCAGGACAGAAG 1080

Qy   1197 GCTAGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCA 1256
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 GCTAGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCA 1140

Qy   1257 GCAGCCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGA 1316
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 GCAGCCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGA 1200

Qy   1317 CACTCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATG 1376
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 CACTCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATG 1260

Qy   1377 GACCATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGG 1436
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 GACCATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGG 1320

Qy   1437 GGAAAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGCT 1496
          |||||||||||||||||||||||||||||||||||||||||||||||| ||| ||||||
Db    1321 GGAAAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGACGCCGACTGCT 1380

Qy   1497 CCCCCAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAG 1556
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 CCCCCAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAG 1440

Qy   1557 AGAGAAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTG 1616
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 AGAGAAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTG 1500

Qy   1617 CACCTCAATTCTCTCTTTGGAGGAGACCAGTAG 1649
          ||||||||||||||||||||||||||||||||
Db    1501 CACCTCAATTCTCTCTTTGGAGGAGACCAGTAG 1533
```

Search completed: November 16, 2002, 11:17:22
Job time : 3348.87 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 16, 2002, 06:33:28 ; Search time 263.38 Seconds
(without alignments)
16356.861 Million cell updates/sec

Title: M33262_COPY_1193_3105
Perfect score: 1913
Sequence: 1 GGTACCAGACGGCGTGAGGA.....TCTGGGGATGTCTCTAAATT 1913

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1913	100.0	10006	24	AAD25514	Simian-Human immun
2	1913	100.0	10097	13	AAQ24802	SIVmac239 nef-dele
3	1913	100.0	10279	13	AAQ22487	SIVmac239 proviral
4	1913	100.0	10279	20	AAV81865	Nucleic acid seque
5	1884.8	98.5	10274	24	AAD25516	Simian-Human immun
6	1788.6	93.5	9601	9	AAN80860	Sequence corresp.
7	1411.4	73.8	1533	22	AAH22809	DNA sequence of a
8	1174.4	61.4	10163	18	AAT61085	Full-length HIV-2K
9	1158.6	60.6	9822	10	AAN91774	Entire HIV-2/ST pr
10	1147.4	60.0	4358	24	AAI67895	Nucleotide sequenc
11	1146.6	59.9	9672	13	AAQ20616	ROD HIV-2 isolate
12	1145	59.9	9630	21	AAZ89628	HIV-2 ROD DNA. Hu
13	1143.4	59.8	9643	9	AAN80859	Sequence of entire
14	1142.4	59.7	9360	11	AAQ02830	cDNA to HIV-2 RNA.
15	1132	59.2	9663	21	AAC81938	Packaging vector p
16	1132	59.2	9726	21	AAC81937	Packaging vector p
17	1129.4	59.0	9473	10	AAN92768	HIV-2 variant HIV-
18	1128.4	59.0	9633	9	AAN80890	Sequence of cDNA c
19	1091.4	57.1	9285	24	ABK91608	Modified HIV prote
20	1071.4	56.0	9280	24	ABK91615	Modified HIV prote
21	1058.2	55.3	9693	10	AAN92119	Sequence of clone
22	987.2	51.6	6486	24	ABK91606	Modified HIV prote
23	977	51.1	1566	8	AAN70887	HIV virus gag sequ
24	977	51.1	1566	9	AAN82198	GAGRODN sequence f
25	969	50.7	1790	10	AAN92770	HIV-2 variant HIV-
26	953.6	49.8	1566	24	ABK52307	DNA encoding retro
27	952	49.8	1566	24	ABK52304	cDNA encoding huma
28	925.8	48.4	1790	16	AAQ92469	HIV-D205 partial s
29	679	35.5	9465	21	AAA51008	Simian immunodefic
30	610	31.9	9641	22	AAH77486	Mandrill immunodef
31	608.4	31.8	2500	22	AAH77480	Mandrill immunodef
32	548.6	28.7	9170	11	AAQ02829	DNA complementary
33	548.6	28.7	9170	22	AAF27365	SIV-derived retrov
34	512.6	26.8	9213	7	AAN60288	Sequence of the HT
35	512.6	26.8	9213	20	AAV81866	Nucleic acid seque
36	512.2	26.8	9213	20	AAV74271	HTLV-III genomic D
37	509.6	26.6	8932	22	AAH76385	Nucleotide sequenc
38	509.6	26.6	8933	22	AAF60542	Plasmid BH10 RNA t
39	509.6	26.6	8933	22	AAF60543	IAC-Ascr pseudo ta
40	509.6	26.6	8933	22	AAF60548	IAC-Bscr pseudo ta
41	509.6	26.6	9425	14	AAQ45919	HTLV-III DNA (Clon
42	509.6	26.6	9427	14	AAQ45920	HTLV-III DNA (Clon
43	509.6	26.6	9749	14	AAQ45922	HTLV-III DNA (Clon
44	503.4	26.3	9745	7	AAN60240	HTLV-III virus (HI
45	501.6	26.2	8966	21	AAF57923	HIV-1 non-subtype

ALIGNMENTS

RESULT 1
AAD25514
ID AAD25514 standard; DNA; 10006 BP.
XX
AC AAD25514;
XX
DT 26-MAR-2002 (first entry)
XX
DE Simian-Human immunodeficiency virus clone SHIV-HXBcP 3.2 complete genome.
XX
KW HIV; human immunodeficiency virus; simian virus; cytostatic; cancer;
KW immunosuppressive; virucide; antibacterial; fungicide; protozoacide;
KW antirheumatic; antiinflammatory; antiarthritic; rheumatoid arthritis;
KW neuroprotective; multiple sclerosis; immune response; vasotropic;
KW vaccine; gene therapy; autoimmune disease; vasculitis; ds.
XX
OS Chimeric - Human immunodeficiency virus.
OS Chimeric - Simian virus.
XX
PN WO200176643-A1.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-US11372.
XX
PR 07-APR-2000; 2000US-195680P.
XX
PA (BAYU) BAYLOR COLLEGE MEDICINE.
XX
PI Orson FM, Kinsey BM, Bhogal BS;
XX
DR WPI; 2002-066308/09.

XX
PT Composition for oral delivery of vaccines, comprises expression vector
PT containing antigenic genomic sequence, bound to aggregated
PT protein-polycationic polymer conjugate or suspension -
XX
PS Disclosure; Page 76-79; 145pp; English.
XX
CC The invention relates to a composition comprising an expression vector
CC bound to an aggregated protein-polycationic polymer conjugate or
CC suspension. The expression vector contains a promoter polynucleotide
CC sequence operatively linked to a polynucleotide sequence encoding an
CC antigen which is a fragment of a gene or genome associated with an
CC infectious disease, cancer and autoimmune disease such as rheumatoid
CC arthritis, vasculitis, and multiple sclerosis; pathogenic genomes
CC consisting of bacterium, fungus, protozoa and virus such as human
CC immunodeficiency virus (HIV), herpes simplex virus (HSV), hepatitis C
CC virus (HCV), influenza and respiratory syncytial virus (RSV), and
CC optionally comprising a nucleotide sequence encoding a cytokine (or a
CC cytokine expression vector), is useful for inducing an immune response
CC (systemic and/or mucosal) in an organism. The cytokine expression vector
CC contains a sequence for granulocyte macrophage-colony stimulating factor
CC (GM-CSF) or interleukin-12 (IL-12). The polynucleotide sequences encoding
CC the antigen and the cytokine are under transcriptional control of same or
CC different promoter polynucleotide sequences. The expression vector, as a
CC DNA vaccine is useful for treating a condition in an organism. The
CC present sequence is simian-human immunodeficiency virus clone
CC SHIV-HXBcP 3.2 complete genomic DNA, related to the invention.
XX
SQ Sequence 10006 BP; 3459 A; 1824 C; 2505 G; 2218 T; 0 other;

Query Match 100.0%; Score 1913; DB 24; Length 10006;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1913; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 GGTACCAGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGCAAC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     420 GGTACCAGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGCAAC 479

Qy      61 ACAAAAAAGAAATAGCTGTCTTTTATCCAGGAAGGGGTAATAAGATAGAGTGGGAGATGG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     480 ACAAAAAAGAAATAGCTGTCTTTTATCCAGGAAGGGGTAATAAGATAGAGTGGGAGATGG 539

Qy     121 GCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCTAC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     540 GCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCTAC 599

Qy     181 GACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTAG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     600 GACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTAG 659

Qy     241 ATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTTT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     660 ATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTTT 719

Qy     301 CGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACTG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     720 CGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACTG 779

Qy     361 TCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAAC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     780 TCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAAC 839

Qy     421 AGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAAACAA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     840 AGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAAACAA 899

Qy     481 GTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGGTG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     900 GTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGGTG 959

Qy     541 GTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGATAG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     960 GTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGATAG 1019

Qy     601 AGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGTTGCA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1020 AGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGTTGCA 1079

Qy     661 CCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCAGA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1080 CCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCAGA 1139

Qy     721 TTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACCAG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1140 TTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACCAG 1199

Qy     781 CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACTAGTT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1200 CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACTAGTT 1259

Qy     841 CAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGCAACA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1260 CAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGCAACA 1319

Qy     901 TTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCAACAA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1320 TTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCAACAA 1379

Qy     961 ACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1380 ACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCT 1439

Qy    1021 ACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACAC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1440 ACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACAC 1499

Qy    1081 TGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATC 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1500 TGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATC 1559

Qy    1141 CCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCTA 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1560 CCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCTA 1619

Qy    1201 GATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCAGCAG 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1620 GATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCAGCAG 1679

Qy    1261 CCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGACACT 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1680 CCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGACACT 1739

Qy    1321 CTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATGGACC 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1740 CTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATGGACC 1799

Qy    1381 ATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGGGGAA 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1800 ATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGGGGAA 1859

Qy    1441 AGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGCTCCCC 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1860 AGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGCTCCCC 1919

Qy    1501 CAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAGAGAG 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1920 CAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAGAGAG 1979

Qy    1561 AAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCACC 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1980 AAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCACC 2039

Qy    1621 TCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAGA 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2040 TCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAGA 2099
```

```
Qy    1681 AGTATTACTGGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCACA 1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2100 AGTATTACTGGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCACA 2159

Qy    1741 TTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAAA 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2160 TTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAAA 2219

Qy    1801 TGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCCC 1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2220 TGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCCC 2279

Qy    1861 GATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAATT 1913
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2280 GATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAATT 2332


RESULT 2
AAQ24802
ID   AAQ24802 standard; DNA; 10097 BP.
XX
AC   AAQ24802;
XX
DT   06-JUL-1992  (first entry)
XX
DE   SIVmac239 nef-deletion.
XX
KW   Macaque; monkey; polymerase chain reaction;
KW   PCR; site-directed mutagenesis; retrovirus; null mutation; ss.
XX
OS   Simian immunodeficiency virus.
XX
FH   Key             Location/Qualifiers
FT   repeat_region   1..818
FT                   /*tag= a
FT                   /rpt_type= TERMINAL
FT                   /note= "i.e. 5' LTR"
FT   repeat_unit     1..517
FT                   /*tag= b
FT                   /rpt_type= OTHER
FT                   /note= "U3"
FT   repeat_unit     518..600
FT                   /*tag= c
FT                   /rpt_type= OTHER
FT                   /note= "R"
FT   repeat_unit     601..818
FT                   /*tag= d
FT                   /rpt_type= OTHER
FT                   /note= "U5"
FT   primer_bind     822..849
FT                   /*tag= e
FT                   /standard_name= tRNA_PBS
FT   CDS             1053..2585
FT                   /*tag= f
FT                   /product= gag
FT   CDS             2228..5410
FT                   /*tag= g
FT                   /product= pol
FT   CDS             5340..5984
FT                   /*tag= h
FT                   /product= vif
FT   CDS             5812..6150
FT                   /*tag= i
FT                   /product= vpx
FT   CDS             6051..6456
FT                   /*tag= j
FT                   /product= vpr
FT   exon            6302..6597
FT                   /*tag= k
FT                   /product= tat
FT                   /note= "full-length product obtained by splicing"
FT   exon            6528..6597
FT                   /*tag= l
FT                   /product= rev
FT                   /note= "full-length product obtained by splicing"
FT   CDS             6604..9243
FT                   /*tag= m
FT                   /product= env
FT   exon            8803..8902
FT                   /*tag= n
FT                   /product= tat
FT                   /note= "see above"
FT   exon            8803..9059
FT                   /*tag= o
FT                   /product= rev
FT                   /note= "see above"
FT   CDS             9077..9686
FT                   /*tag= p
FT                   /product= nef
FT   repeat_region   9280..10097
FT                   /*tag= q
FT                   /rpt_type= TERMINAL
FT                   /note= "i.e. 3' LTR"
FT   misc_signal     415..424
FT                   /*tag= r
FT                   /standard_name= NF_Kappa_B
FT   GC_signal       429..438
FT                   /*tag= s
FT                   /standard_name= Sp1_binding_site
FT   GC_signal       440..449
FT                   /*tag= t
FT                   /standard_name= Sp1_binding_site
FT   GC_signal       451..460
FT                   /*tag= u
FT                   /standard_name= Sp1_binding_site
FT   GC_signal       462..471
FT                   /*tag= v
FT                   /standard_name= Sp1_binding_site
FT   TATA_signal     488..494
FT                   /*tag= w
FT   polyA_signal    9950..9955
FT                   /*tag= x
XX
PN   WO9200987-A.
XX
PD   23-JAN-1992.
XX
PF   10-JUL-1991;   91WO-US04884.
XX
PR   12-JUL-1990;   90US-0551945.
XX
PA   (HARD ) HARVARD COLLEGE.
XX
PI   Desrosiers RC;
XX
DR   WPI; 1992-056816/07.
DR   P-PSDB; AAR22365-R22371, AAR24126-7.
XX
PT   Primate lentivirus vaccine protecting against AIDS - and primate
PT   lentiviruses and their DNA clones contg. null mutations, useful for
PT   producing vaccine
XX
PS   Claim 1; Fig 1; 51pp; English.
XX
CC   Parental virus SIVmac239 was isolated from a macaque monkey (see
CC   AAQ22487). An SstI fragment of p239SpE3' contg. the C-terminus of
CC   the nef gene was cloned into M13 and subjected to site-directed
CC   mutagenesis. A 73mer primer was used which was complementary to
CC   bases 9215 through 9250 and 9433 through 9469. Since bases 9251
CC   through 9432 were not included in the primer, their complement was
CC   not included in the newly synthesised negative strand. Successful
CC   deletion was confirmed by DNA sequencing. The fragment contg. the
CC   deletion was cloned into p239SpE3' to yield p239S-E3'(nef-deletion)
CC   which was digested with SphI and ligated to SphI-cut p239SpSp5'.
CC   The ligation product was used to directly transfect cultured cells.
CC   See also AAQ21075-8.
```

```
XX
SQ   Sequence 10097 BP; 3387 A; 1911 C; 2527 G; 2272 T; 0 other;

  Query Match             100.0%;  Score 1913;  DB 13;  Length 10097;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1913;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GGTACCAGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGCAAC 60
Db     937 GGTACCAGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGCAAC 996

Qy      61 ACAAAAAAGAAATAGCTGTCTTTTATCCAGGAAGGGGTAATAAGATAGAGTGGGAGATGG 120
Db     997 ACAAAAAAGAAATAGCTGTCTTTTATCCAGGAAGGGGTAATAAGATAGAGTGGGAGATGG 1056

Qy     121 GCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCTAC 180
Db    1057 GCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCTAC 1116

Qy     181 GACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTAG 240
Db    1117 GACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTAG 1176

Qy     241 ATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTTT 300
Db    1177 ATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTTT 1236

Qy     301 CGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACTG 360
Db    1237 CGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACTG 1296

Qy     361 TCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAAC 420
Db    1297 TCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAAC 1356

Qy     421 AGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAAACAA 480
Db    1357 AGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAAACAA 1416

Qy     481 GTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGGTG 540
Db    1417 GTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGGTG 1476

Qy     541 GTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGATAG 600
Db    1477 GTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGATAG 1536

Qy     601 AGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGTTGCA 660
Db    1537 AGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGTTGCA 1596

Qy     661 CCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCAGA 720
Db    1597 CCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCAGA 1656

Qy     721 TTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACCAG 780
Db    1657 TTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACCAG 1716

Qy     781 CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACTAGTT 840
Db    1717 CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACTAGTT 1776

Qy     841 CAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGCAACA 900
Db    1777 CAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGCAACA 1836

Qy     901 TTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCAACAA 960
Db    1837 TTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCAACAA 1896

Qy     961 ACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCT 1020
Db    1897 ACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCT 1956

Qy    1021 ACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACAC 1080
Db    1957 ACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACAC 2016

Qy    1081 TGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATC 1140
Db    2017 TGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATC 2076

Qy    1141 CCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCTA 1200
Db    2077 CCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCTA 2136

Qy    1201 GATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCAGCAG 1260
Db    2137 GATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCAGCAG 2196

Qy    1261 CCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGACACT 1320
Db    2197 CCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGACACT 2256

Qy    1321 CTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATGGACC 1380
Db    2257 CTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATGGACC 2316

Qy    1381 ATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGGGGAA 1440
Db    2317 ATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGGGGAA 2376

Qy    1441 AGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGCTCCCC 1500
Db    2377 AGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGCTCCCC 2436

Qy    1501 CAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAGAGAG 1560
Db    2437 CAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAGAGAG 2496

Qy    1561 AAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCACC 1620
Db    2497 AAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCACC 2556

Qy    1621 TCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAGA 1680
Db    2557 TCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAGA 2616

Qy    1681 AGTATTACTGGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCACA 1740
Db    2617 AGTATTACTGGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCACA 2676

Qy    1741 TTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAAA 1800
Db    2677 TTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAAA 2736

Qy    1801 TGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCCC 1860
Db    2737 TGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCCC 2796

Qy    1861 GATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAATT 1913
Db    2797 GATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAATT 2849


RESULT 3
AAQ22487
ID   AAQ22487 standard; DNA; 10279 BP.
XX
AC   AAQ22487;
XX
DT   06-JUL-1992  (first entry)
XX
DE   SIVmac239 proviral genome.
XX
```

```
KW   Macaque; monkey; mac239; polymerase chain reaction;
KW   PCR; site-directed mutagenesis; retrovirus; ss.
XX
OS   Simian immunodeficiency virus.
XX
FH   Key             Location/Qualifiers
FT   repeat_region   1..818
FT                   /*tag= a
FT                   /rpt_type= TERMINAL
FT                   /note= "i.e. 5' LTR"
FT   repeat_unit     1..517
FT                   /*tag= b
FT                   /rpt_type= OTHER
FT                   /note= "U3"
FT   repeat_unit     518..600
FT                   /*tag= c
FT                   /rpt_type= OTHER
FT                   /note= "R"
FT   repeat_unit     601..818
FT                   /*tag= d
FT                   /rpt_type= OTHER
FT                   /note= "U5"
FT   primer_bind     822..849
FT                   /*tag= e
FT                   /standard_name= tRNA_PBS
FT   CDS             1053..2585
FT                   /*tag= f
FT                   /product= gag
FT   CDS             2228..5410
FT                   /*tag= g
FT                   /product= pol
FT   CDS             5340..5984
FT                   /*tag= h
FT                   /product= vif
FT   CDS             5812..6150
FT                   /*tag= i
FT                   /product= vpx
FT   CDS             6051..6456
FT                   /*tag= j
FT                   /product= vpr
FT   exon            6302..6597
FT                   /*tag= k
FT                   /product= tat
FT                   /note= "full-length product obtained by splicing"
FT   exon            6528..6597
FT                   /*tag= l
FT                   /product= rev
FT                   /note= "full-length product obtained by splicing"
FT   CDS             6604..9243
FT                   /*tag= m
FT                   /product= env
FT   exon            8803..8902
FT                   /*tag= n
FT                   /product= tat
FT                   /note= "see above"
FT   exon            8803..9059
FT                   /*tag= o
FT                   /product= rev
FT                   /note= "see above"
FT   CDS             9077..9868
FT                   /*tag= p
FT                   /product= nef
FT   repeat_region   9462..10279
FT                   /*tag= q
FT                   /rpt_type= TERMINAL
FT                   /note= "i.e. 3' LTR"
FT   misc_signal     415..424
FT                   /*tag= r
FT                   /standard_name= NF_Kappa_B
FT   GC_signal       429..438
FT                   /*tag= s
FT                   /standard_name= Sp1_binding_site
FT   GC_signal       440..449
FT                   /*tag= t
FT                   /standard_name= Sp1_binding_site
FT   GC_signal       451..460
FT                   /*tag= u
FT                   /standard_name= Sp1_binding_site
FT   GC_signal       462..471
FT                   /*tag= v
FT                   /standard_name= Sp1_binding_site
FT   TATA_signal     488..494
FT                   /*tag= w
FT   polyA_signal    10132..10137
FT                   /*tag= x
XX
PN   WO9200987-A.
XX
PD   23-JAN-1992.
XX
PF   10-JUL-1991;   91WO-US04884.
XX
PR   12-JUL-1990;   90US-0551945.
XX
PA   (HARD ) HARVARD COLLEGE.
XX
PI   Desrosiers RC;
XX
DR   WPI; 1992-056816/07.
DR   P-PSDB; AAR22365-R22372, AAR24126.
XX
PT   Primate lentivirus vaccine protecting against AIDS - and primate
PT   lentiviruses and their DNA clones contg. null mutations, useful for
PT   producing vaccine
XX
PS   Disclosure; Fig 1; 51pp; English.
XX
CC   Cell-free serum samples from a macaque monkey exhibiting symptoms
CC   characteristic of SIV infection were co-cultivated with HuT-78
CC   cells. Infectious SIVmac239 virus was identified in the cell
CC   supernatant. Total cell DNA was prepared from SIVmac239-infected
CC   cells and digested with EcoRI. An EMBL-4 library was constructed
CC   from 10-20kb EcoRI fragments (EcoRI is a non-cutter of SIVmac239).
CC   The library was screened with pK2 BamA as probe and a full-length
CC   molecular clone was isolated and sequenced. Then, EMBL-SIVmac239
CC   was digested with SphI and a 6706bp fragment, contg. the SphI site
CC   in the left flanking cellular DNA sequence to viral nucleotide no.
CC   6451, was inserted in vector pBS(+) to produce subclone p239SpSp5'.
CC   In a separate reaction, EMBL-SIVmac239 was digested with EcoRI and
CC   SphI. A 6361bp fragment from viral nucleotide 6452 to the EcoRI
CC   site in the right flanking cellular sequence, was inserted in
CC   pBS(-) to produce subclone p239SpE3'. These subclones were used to
CC   generate the full-length genomic sequence and to produce the
CC   preferred null-mutations of the invention. See Q24802 for
CC   nef-deletion mutant. See also Q21075-8.
XX
SQ   Sequence 10279 BP; 3465 A; 1936 C; 2569 G; 2309 T; 0 other;

  Query Match            100.0%;  Score 1913;  DB 13;  Length 10279;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1913;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGTACCAGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGCAAC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    937 GGTACCAGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGCAAC 996

Qy     61 ACAAAAAAGAAATAGCTGTCTTTTATCCAGGAAGGGGTAATAAGATAGAGTGGGAGATGG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    997 ACAAAAAAGAAATAGCTGTCTTTTATCCAGGAAGGGGTAATAAGATAGAGTGGGAGATGG 1056

Qy    121 GCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCTAC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1057 GCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCTAC 1116

Qy    181 GACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTAG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1117 GACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTAG 1176

Qy    241 ATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTTT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1177 ATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTTT 1236

Qy    301 CGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACTG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1237 CGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACTG 1296

Qy    361 TCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAAC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1297 TCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAAC 1356

Qy    421 AGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAAACAA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1357 AGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAAACAA 1416

Qy    481 GTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGGTG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1417 GTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGGTG 1476

Qy    541 GTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGATAG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1477 GTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGATAG 1536

Qy    601 AGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGTTGCA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1537 AGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGTTGCA 1596

Qy    661 CCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCAGA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1597 CCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCAGA 1656

Qy    721 TTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACCAG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1657 TTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACCAG 1716

Qy    781 CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACTAGTT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1717 CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACTAGTT 1776

Qy    841 CAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGCAACA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1777 CAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGCAACA 1836

Qy    901 TTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCAACAA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1837 TTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCAACAA 1896

Qy    961 ACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1897 ACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCT 1956

Qy   1021 ACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACAC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1957 ACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACAC 2016

Qy   1081 TGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATC 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2017 TGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATC 2076

Qy   1141 CCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCTA 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2077 CCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCTA 2136

Qy   1201 GATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCAGCAG 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2137 GATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCAGCAG 2196

Qy   1261 CCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGACACT 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2197 CCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGACACT 2256

Qy   1321 CTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATGGACC 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2257 CTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATGGACC 2316

Qy   1381 ATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGGGGAA 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2317 ATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGGGGAA 2376

Qy   1441 AGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGCTCCCC 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2377 AGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGCTCCCC 2436

Qy   1501 CAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAGAGAG 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2437 CAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAGAGAG 2496

Qy   1561 AAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCACC 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2497 AAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCACC 2556

Qy   1621 TCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAGA 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2557 TCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAGA 2616

Qy   1681 AGTATTACTGGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCACA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2617 AGTATTACTGGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCACA 2676

Qy   1741 TTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAAA 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2677 TTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAAA 2736

Qy   1801 TGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCCC 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2737 TGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCCC 2796

Qy   1861 GATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAATT 1913
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2797 GATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAATT 2849
```

```
RESULT 4
AAV81865
ID   AAV81865 standard; DNA; 10279 BP.
XX
AC   AAV81865;
XX
DT   01-JUN-1999  (first entry)
XX
DE   Nucleic acid sequence of SIVmac239 genome.
XX
KW   Antigenic composition; primate; lentivirus; nef gene; vaccine;
KW   infection; AIDS; SIVmac239; ss.
XX
OS   Simian immunodeficiency virus.
XX
PN   US5851813-A.
XX
PD   22-DEC-1998.
XX
PF   27-JAN-1994;   94US-0188583.
XX
PR   27-JAN-1994;   94US-0188583.
PR   12-JUL-1990;   90US-0551945.
PR   09-JUL-1991;   91US-0727494.
XX
PA   (HARD ) HARVARD COLLEGE.
XX
PI   Desrosiers RC;
XX
DR   WPI; 1999-080408/07.
DR   P-PSDB; AAW89313, AAW89314, AAW89315, AAW89316, AAW89317, AAW89318,
```

```
DR   AAW89319, AAW89320, AAW89321.
XX
PT   Lentivirus antigenic compositions - containing lentivirus with nef
PT   gene deletion
XX
PS   Disclosure; Fig 1A-V; 93pp; English.
XX
CC   The invention relates to an antigenic composition comprising an isolated
CC   primate lentivirus whose genome contains an engineered non-revertible
CC   null mutation in the nef gene, or an infectious DNA clone in a carrier.
CC   The antigenic composition is used in vaccines against infection by the
CC   lentivirus,  e.g. AIDS.  The present sequence represents the nucleic
CC   acid sequence of the SIVmac 239 genome.
XX
SQ   Sequence 10279 BP; 3466 A; 1935 C; 2570 G; 2308 T; 0 other;

  Query Match             100.0%;  Score 1913;  DB 20;  Length 10279;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1913;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGTACCAGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGCAAC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    937 GGTACCAGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGCAAC 996

Qy     61 ACAAAAAAGAAATAGCTGTCTTTTATCCAGGAAGGGGTAATAAGATAGAGTGGGAGATGG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    997 ACAAAAAAGAAATAGCTGTCTTTTATCCAGGAAGGGGTAATAAGATAGAGTGGGAGATGG 1056

Qy    121 GCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCTAC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1057 GCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCTAC 1116

Qy    181 GACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTAG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1117 GACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTAG 1176

Qy    241 ATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTTT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1177 ATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTTT 1236

Qy    301 CGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACTG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1237 CGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACTG 1296

Qy    361 TCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAAC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1297 TCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAAC 1356

Qy    421 AGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAAACAA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1357 AGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAAACAA 1416

Qy    481 GTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGGTG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1417 GTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGGTG 1476

Qy    541 GTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGATAG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1477 GTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGATAG 1536

Qy    601 AGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGTTGCA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1537 AGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGTTGCA 1596

Qy    661 CCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCAGA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1597 CCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCAGA 1656

Qy    721 TTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACCAG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1657 TTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACCAG 1716

Qy    781 CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACTAGTT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1717 CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACTAGTT 1776

Qy    841 CAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGCAACA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1777 CAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGCAACA 1836

Qy    901 TTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCAACAA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1837 TTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCAACAA 1896

Qy    961 ACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1897 ACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCT 1956

Qy   1021 ACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACAC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1957 ACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACAC 2016

Qy   1081 TGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATC 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2017 TGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATC 2076

Qy   1141 CCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCTA 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2077 CCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCTA 2136

Qy   1201 GATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCAGCAG 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2137 GATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCAGCAG 2196

Qy   1261 CCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGACACT 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2197 CCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGACACT 2256

Qy   1321 CTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATGGACC 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2257 CTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATGGACC 2316

Qy   1381 ATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGGGGAA 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2317 ATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGGGGAA 2376

Qy   1441 AGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGCTCCCC 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2377 AGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGCTCCCC 2436

Qy   1501 CAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAGAGAG 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2437 CAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAGAGAG 2496

Qy   1561 AAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCACC 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2497 AAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCACC 2556

Qy   1621 TCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAGA 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2557 TCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAGA 2616

Qy   1681 AGTATTACTGGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCACA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2617 AGTATTACTGGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCACA 2676

Qy   1741 TTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAAA 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2677 TTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAAA 2736

Qy   1801 TGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCCC 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2737 TGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCCC 2796

Qy   1861 GATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAATT 1913
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   2797 GATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAATT 2849


RESULT 5
AAD25516
ID   AAD25516 standard; DNA; 10274 BP.
XX
AC   AAD25516;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Simian-Human immunodeficiency virus (SHIV) clone 1A11 complete genome.
XX
KW   HIV; human immunodeficiency virus; simian virus; cytostatic; cancer;
KW   immunosuppressive; virucide; antibacterial; fungicide; protozoacide;
KW   antirheumatic; antiinflammatory; antiarthritic; rheumatoid arthritis;
KW   neuroprotective; multiple sclerosis; immune response; vasotropic;
KW   vaccine; gene therapy; autoimmune disease; vasculitis; ds.
XX
OS   Chimeric - Human immunodeficiency virus.
OS   Chimeric - Simian virus.
XX
PN   WO200176643-A1.
XX
PD   18-OCT-2001.
XX
PF   06-APR-2001; 2001WO-US11372.
XX
PR   07-APR-2000; 2000US-195680P.
XX
PA   (BAYU ) BAYLOR COLLEGE MEDICINE.
XX
PI   Orson FM,  Kinsey BM,  Bhogal BS;
XX
DR   WPI; 2002-066308/09.
XX
PT   Composition for oral delivery of vaccines, comprises expression vector
PT   containing antigenic genomic sequence, bound to aggregated
PT   protein-polycationic polymer conjugate or suspension  -
XX
PS   Disclosure; Page 82-84; 145pp; English.
XX
CC   The invention relates to a composition comprising an expression vector
CC   bound to an aggregated protein-polycationic polymer conjugate or
CC   suspension. The expression vector contains a promoter polynucleotide
CC   sequence operatively linked to a polynucleotide sequence encoding an
CC   antigen which is a fragment of a gene or genome associated with an
CC   infectious disease, cancer and autoimmune disease such as rheumatoid
CC   arthritis, vasculitis, and multiple sclerosis, pathogenic genomes
CC   consisting of bacterium, fungus, protozoa and virus such as human
CC   immunodeficiency virus (HIV), herpes simplex virus (HSV), hepatitis C
CC   virus (HCV), influenza and respiratory syncytial virus (RSV), and
CC   optionally comprising a nucleotide sequence encoding a cytokine (or a
CC   cytokine expression vector), is useful for inducing an immune response
CC   (systemic and/or mucosal) in an organism. The cytokine expression vector
CC   contains a sequence for granulocyte macrophage-colony stimulating factor
CC   (GM-CSF) or interleukin-12 (IL-12). The polynucleotide sequences encoding
CC   the antigen and the cytokine are under transcriptional control of same or
CC   different promoter polynucleotide sequences. The expression vector, as a
CC   DNA vaccine is useful for treating a condition in an organism. The
CC   present sequence is simian-human immunodeficiency virus (SHIV) clone
CC   1A11 complete genomic DNA, related to the invention.
XX
SQ   Sequence 10274 BP; 3475 A; 1931 C; 2559 G; 2309 T; 0 other;

  Query Match             98.5%;  Score 1884.8;  DB 24;  Length 10274;
  Best Local Similarity   99.1%;  Pred. No. 0;
  Matches 1895;  Conservative    0;  Mismatches   17;  Indels    0;  Gaps    0;

Qy       1 GGTACCAGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGCAAC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     936 GGTACCAGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGCAAC 995

Qy      61 ACAAAAAAGAAATAGCTGTCTTTTATCCAGGAAGGGGTAATAAGATAGAGTGGGAGATGG 120
           |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db     996 ACAAAAAAGAAATAGCTGTCTTTTATCCAGGAAGGGATAATAAGATAGAGTGGGAGATGG 1055

Qy     121 GCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCTAC 180
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1056 GCGCGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCTAC 1115

Qy     181 GACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTAG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1116 GACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTAG 1175

Qy     241 ATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTTT 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1176 ATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTTT 1235

Qy     301 CGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACTG 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1236 CGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACTG 1295

Qy     361 TCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAAC 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1296 TCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAAC 1355

Qy     421 AGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAAACAA 480
           ||||||| |||||||||||||||||||||||||||| |||||| ||||||||||||||||
Db    1356 AGATAGTACAGAGACACCTAGTGGTGGAAACAGGAACAGCAGAAATTATGCCAAAAACAA 1415

Qy     481 GTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGGTG 540
           | |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1416 GCAGACCAACTGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGGTG 1475

Qy     541 GTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGATAG 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1476 GTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGATAG 1535

Qy     601 AGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGTTGCA 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1536 AGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGTTGCA 1595

Qy     661 CCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCAGA 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1596 CCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCAGA 1655

Qy     721 TTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACCAG 780
           |||||||||| || ||||||||||||||||||||||||||||||||||||||||||||||
Db    1656 TTATCAGAGACATCATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACCAG 1715

Qy     781 CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACTAGTT 840
           ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    1716 CTCCACAACAAGGACAGCTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACTAGTT 1775

Qy     841 CAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGCAACA 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1776 CAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGCAACA 1835

Qy     901 TTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCAACAA 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1836 TTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCAACAA 1895

Qy     961 ACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCT 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1896 ACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCT 1955

Qy    1021 ACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACAC 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1956 ACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACAC 2015

Qy    1081 TGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATC 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2016 TGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATC 2075

Qy    1141 CCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCTA 1200
```

```
Db    2076 CCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCAGGACAGAAGGCTA 2135

Qy    1201 GATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCAGCAG 1260
Db    2136 GATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCAGCAG 2195

Qy    1261 CCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGACACT 1320
Db    2196 CCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGACACT 2255

Qy    1321 CTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATGGACC 1380
Db    2256 CTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATGGACC 2315

Qy    1381 ATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGGGGAA 1440
Db    2316 ATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGGGGAA 2375

Qy    1441 AGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGCTCCCC 1500
Db    2376 AGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGACGCCAACTGCTCCCC 2435

Qy    1501 CAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAGAGAG 1560
Db    2436 CAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAGAGAG 2495

Qy    1561 AAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCACC 1620
Db    2496 AAAAACAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGATGACAGAGGATTTGCTGCACC 2555

Qy    1621 TCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAGA 1680
Db    2556 TCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAGA 2615

Qy    1681 AGTATTACTGGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCACA 1740
Db    2616 AGTATTACTGGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCACA 2675

Qy    1741 TTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAAA 1800
Db    2676 TTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAAA 2735

Qy    1801 TGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCCC 1860
Db    2736 TGTAGAAATAGAAGTTTTAGGCAAAAGAATTAGAGGGACAATCATGACAGGGGACACCCC 2795

Qy    1861 GATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAAT 1912
Db    2796 GATTAACATTTTTGGTAGAAATTTACTAACAGCTCTGGGGATGTCTCTAAAT 2847


RESULT 6
AAN80860
ID   AAN80860 standard; cDNA; 9601 BP.
XX
AC   AAN80860;
XX
DT   15-NOV-1990  (first entry)
XX
DE   Sequence corresp. to the genome of SIV isolate STLV-III Mac 142-83
DE   in clone SIV-1.
XX
KW   pSIV-1.1; pSIV-1.2; STLV-III Mac 142-83; ss.
XX
OS   Simian immunodeficiency virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             551..2068
FT                   /*tag=  a
FT                   /gene="gagd"
FT   CDS             1726..4893
FT                   /*tag=  b
FT                   /gene="pol"
FT   CDS             4826..5467
FT                   /*tag=  c
FT                   /gene="Q"
FT   CDS             5298..5633
FT                   /*tag=  d
FT                   /gene="X"
FT   CDS             5637..5939
FT                   /*tag=  e
FT                   /gene="R"
FT   CDS             5788..6084
FT                   /*tag=  f
FT                   /gene="tat1"
FT   CDS             6014..6130
FT                   /*tag=  g
FT                   /gene="art1"
FT   CDS             6090..8297
FT                   /*tag=  h
FT                   /gene="env"
FT   CDS             8298..8732
FT                   /*tag=  i
FT                   /gene="env ctd."
FT   CDS             8296..8391
FT                   /*tag=  j
FT                   /gene="tat2"
FT   CDS             8294..8548
FT                   /*tag=  k
FT                   /gene="art2"
FT   CDS             8569..9354
FT                   /*tag=  l
FT                   /gene="F"
XX
PN   WO8805440-A.
XX
PD   28-JUL-1988.
XX
PF   15-JAN-1988;   88WO-FR00025.
XX
PR   15-APR-1987;   87FR-0005398.
XX
PA   (INSP ) INST PASTEUR.
PA   (ALIZ/) ALIZON M.
XX
PI   Alizon M,  Montagnier L,  Guetard D,  Clavfl F,  Sonigo P,  Guyader M;
PI   Tiollais P,  Chakrabarti L,  Desrosiers R;
XX
DR   WPI; 1988-220290/31.
DR   P-PSDB; AAP80802, AAP81783, AAP81784, AAP81785, AAP81691, AAP81756,
DR   AAP81757, AAP81758, AAP81759, AAP81760, AAP81761, AAP81762.
XX
PT   New peptide(s) with immunological properties of HIV-2 envelope protein -
PT   having the structure of simian immune deficiency virus proteins,
PT   useful in diagnosis and of vaccine components
XX
PS   Claim 23; Fig 1B; 86pp; French.
XX
CC   AAN80860 is contained in lambda SIV-1 which was deposited on 15/4/87 at
CC   the CNCM under numbers I-658 (pSIV-1.1) and I-659 (pSIV-1.2). It is the
CC   cDNA to SIV-1 genomic RNA. Recombinant DNA contg. all or part of the
CC   cDNA from this sequence inserted into a vector are claimed. Labelled
CC   fragments of cDNA can be used as hybridisation probes. New peptides which
CC   have immunological properties in common with those of the peptide
CC   skeleton of the envelope protein of HIV-2 and also have a peptide
CC   structure in common with that of SIV-1 glycoprotein are claimed.
CC   Antigenic and immunogenic conjugates contg. the peptides and a kit to
CC   detect HIV-2 in biological fluids are new. The peptides are useful for in
CC   vitro diagnosis of HIV-2 infection and some of them can be used as
CC   components of immunogens and vaccines against HIV. Antibodies raised
CC   against them can be used for treatment of AIDS.
XX
SQ   Sequence 9601 BP; 3277 A; 1809 C; 2378 G; 2136 T; 1 other;

  Query Match             93.5%;  Score 1788.6;  DB 9;  Length 9601;
```

```
Best Local Similarity   96.8%;  Pred. No. 0;
Matches 1851;  Conservative    0;  Mismatches   49;  Indels   13;  Gaps    2;

Qy       1 GGTACCAGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCC-GGTTGCAGGTAAGTGCAA 59
Db     435 GGTACCAGACGGCGTGAGGAGCGGGGAGAGAAGAGGCCTCCTGGTTGCAGGTAAGTGCAA 494

Qy      60 CACAAAAAAGAAATAGCTGTCTTTTATCCAGGAAGGGGTAATAAGATAGAGTGGGAGATG 119
Db     495 CACAAAAAGGAAATAGCTGTCTTTTATCCAGGAAGGGATAATAAGATAGAGTGGGAGATG 554

Qy     120 GGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCTA 179
Db     555 GGCGCGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGACTA 614

Qy     180 CGACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTA 239
Db     615 CGACCCGGCGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTA 674

Qy     240 GATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTT 299
Db     675 GATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTT 734

Qy     300 TCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACT 359
Db     735 TCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACT 794

Qy     360 GTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAA 419
Db     795 GTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAA 854

Qy     420 CAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAAACA 479
Db     855 CAGATAGTGCAGAGACACCTAGTGATGGAAACAGGAACAGCAGAAACTATGCCAAAAACA 914

Qy     480 AGTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGGT 539
Db     915 AGTAGACCAACAGCACCATTTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGGT 974

Qy     540 GGTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGATA 599
Db     975 GGTAACTATACCCACCTACCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTAATA 1034

Qy     600 GAGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGTTGC 659
Db    1035 GAGGAGAAGAAATTTGGAGCAGAAGTAGTGTCAGGATTTCAGGCACTGTCAGAAGGCTGC 1094

Qy     660 ACCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCAG 719
Db    1095 CTCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCAG 1154

Qy     720 ATTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACCA 779
Db    1155 ATCATCAGAGATATTATAAATGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACAA 1214

Qy     780 GCTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACTAGT 839
Db    1215 GCTCCACAACAAGGACAGCTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACTAGT 1274

Qy     840 TCAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGCAAC 899
Db    1275 ACAGTAGAAGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGCAAC 1334

Qy     900 ATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCAACA 959
Db    1335 ATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCAACA 1394

Qy     960 AACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTC 1019
Db    1395 AACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTC 1454

Qy    1020 TACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACA 1079
Db    1455 TACAAAAGTTTAAGAGCAGAACAAACAGATCCAGCAGTAAAGAATTGGATGACTCAAACA 1514

Qy    1080 CTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAAT 1139
Db    1515 CTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTACGAAT 1574

Qy    1140 CCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCT 1199
Db    1575 CCCACCCTAGAAGAAATGCTGACGGCCTGTCAAGGAGTAGGGGGGCCAGGACAGAAGGCT 1634

Qy    1200 AGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCAGCA 1259
Db    1635 AGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGCGCCAATCCCTTTTGCAGCA 1694

Qy    1260 GCCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGACAC 1319
Db    1695 GCCCAACAGAAGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAGGAAGGACAC 1754

Qy    1320 TCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATGGAC 1379
Db    1755 TCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATGGAC 1814

Qy    1380 CATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGGGGA 1439
Db    1815 CATGTTATGGCCAAATGCCCAAACAGACAGGCGGGTTTTTTAGGCCTTGGCCCTTGGGGA 1874

Qy    1440 AAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGCTCCC 1499
Db    1875 AAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGACGCCAACTGCTCCC 1934

Qy    1500 CCAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAGAGA 1559
Db    1935 CCAGAAGAACCAGCTGTGGATCTGCTAAAGAACTACATGCACTTGGGCAAGC-------- 1986

Qy    1560 GAAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCAC 1619
Db    1987 ----AGCAGAGAGAAAGCAGAGGGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCAC 2042

Qy    1620 CTCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAG 1679
Db    2043 CTCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAG 2102

Qy    1680 AAGTATTACTGGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCAC 1739
Db    2103 AAGTATTATTAGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCAC 2162

Qy    1740 ATTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAA 1799
Db    2163 ATTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAA 2222

Qy    1800 ATGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCC 1859
Db    2223 ATGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCC 2282

Qy    1860 CGATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAAT 1912
Db    2283 CGATTAACATTTTTGGTAGAAATTTACTAACAGCTCTGGGGATGTCTCTAAAT 2335
```

RESULT 7
AAH22809
ID   AAH22809 standard; DNA; 1533 BP.
XX
AC   AAH22809;
XX
DT   07-SEP-2001 (first entry)
XX
DE   DNA sequence of a mutated SIV-1 gag gene.
XX
KW   HIV-1; gag gene; pol gene; SIV-1; env gene; mutation; immunity; AIDS;
KW   gene therapy; immune therapy; immuneprophylaxis; vaccine; retroviral;
KW   lentiviral; anti-HIV; ds.
XX

```
OS   Simian immunodeficiency virus.
XX
PN   WO200146408-A2.
XX
PD   28-JUN-2001.
XX
PF   22-DEC-2000; 2000WO-US34985.
XX
PR   23-DEC-1999;   99US-0173036.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Pavlakis GN;
XX
DR   WPI; 2001-408642/43.
XX
PT   Novel nucleic acid construct containing HIV-1 gag/pol and simian
PT   immunodeficiency virus gag or env genes which have been mutated to
PT   remove inhibitory/instability sequences, useful as components of
PT   anti-HIV vaccines  -
XX
PS   Claim 3; Fig 3; 95pp; English.
XX
CC   The invention provides nucleic acid constructs containing HIV-1 gag/pol
CC   and SIV gag or SIV env genes which have been mutated to remove or
CC   reduce inhibitory/instability sequences. Pharmaceutical compositions
CC   comprising the nucleic acid construct and a carrier is useful for
CC   inducing antibodies or cytotoxic T lymphocytes in a mammal. Vaccine
CC   compositions are useful for inducing immunity against HIV-1 infection in
CC   a human. The constructs are useful for producing infectious viral
CC   particles. The vectors, host cells and vector systems comprising the
CC   constructs are useful in gene therapy, immune therapy, immuneprophylaxis,
CC   vaccine or in genetic therapy after expression. The complete removal of
CC   the inhibitory/instability sequence elements from the lentiviral genome
CC   would result in constitutive expression. This could prevent the virus
CC   from establishing a latent infection and escaping immune system
CC   surveillance. The constructs are also useful for in vivo or in vitro gene
CC   replacement. Mutated genes can be expressed in the native host cell or
CC   organism or in a different cell or organism. The mutated genes can be
CC   introduced into a vector such as a plasmid, cosmid, minichromosome or
CC   virus and inserted into a host cell or organism. Host cells containing
CC   these mutated genes can be used to express large amounts of the protein
CC   useful in enzyme preparations, pharmaceuticals, diagnostic reagents,
CC   vaccines and therapeutics. The constructs and/or proteins made by using
CC   constructs encoding the mutated gag, env, and pol genes could be used
CC   in the production of diagnostic reagents, and therapies for AIDS and
CC   AIDS related diseases. Further, the gag constructs could be used in
CC   transdominant inhibition of HIV expression after the introduction into
CC   humans. The vectors would be introduced into humans, resulting in the
CC   inhibition and of immunostimulation by the produced gag protein. In
CC   addition, the gag constructs could be used in the generation of new
CC   retroviral vectors based on the expression of lentiviral gag proteins.
CC   The present sequence represents the DNA sequence of a mutated SIV-1
CC   gag gene.
XX
SQ   Sequence 1533 BP; 494 A; 354 C; 410 G; 275 T; 0 other;

  Query Match             73.8%;  Score 1411.4;  DB 22;  Length 1533;
  Best Local Similarity   95.0%;  Pred. No. 0;
  Matches 1457;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy     117 ATGGGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGG 176
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGGGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGG 60

Qy     177 CTACGACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAA 236
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CTACGACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAA 120

Qy     237 TTAGATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATA 296
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TTAGATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATA 180

Qy     297 CTTTCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAAT 356
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CTTTCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAAT 240

Qy     357 ACTGTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCA 416
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 ACTGTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCA 300

Qy     417 AAACAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAA 476
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AAACAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACCACCGAAACCATGCCGAAG 360

Qy     477 ACAAGTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATA 536
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ACCTCTCGACCAACAGCACCATCTAGCGGCAGAGGAGGAAACTACCCAGTACAGCAGATC 420

Qy     537 GGTGGTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTG 596
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GGTGGCAACTACGTCCACCTGCCACTGTCCCCGAGAACCCTGAACGCTTGGGTCAAGCTG 480

Qy     597 ATAGAGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGT 656
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ATCGAGGAGAAGAAGTTCGGAGCAGAAGTAGTGCCAGGATTCCAGGCACTGTCAGAAGGT 540

Qy     657 TGCACCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATG 716
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 TGCACCCCCTACGACATCAACCAGATGCTGAACTGCGTTGGAGACCATCAGGCGGCTATG 600

Qy     717 CAGATTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAA 776
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 CAGATCATCCGTGACATCATCAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAA 660

Qy     777 CCAGCTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACT 836
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 CCAGCTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGACATCGCAGGAACCACC 720

Qy     837 AGTTCAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGC 896
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 TCCTCAGTTGACGAACAGATCCAGTGGATGTACCGTCAGCAGAACCCGATCCCAGTAGGC 780

Qy     897 AACATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCA 956
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 AACATCTACCGTCGATGGATCCAGCTGGGTCTGCAGAAATGCGTCCGTATGTACAACCCG 840

Qy     957 ACAAACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGG 1016
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 ACCAACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGG 900

Qy    1017 TTCTACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAA 1076
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 TTCTACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAA 960

Qy    1077 ACACTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTG 1136
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 ACACTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTG 1020

Qy    1137 AATCCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAG 1196
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 AATCCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAG 1080

Qy    1197 GCTAGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCA 1256
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 GCTAGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCA 1140

Qy    1257 GCAGCCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGA 1316
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 GCAGCCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGA 1200

Qy    1317 CACTCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATG 1376
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 CACTCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATG 1260

Qy    1377 GACCATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGG 1436
```

```
Db    1261 GACCATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGG 1320

Qy    1437 GGAAAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGCT 1496

Db    1321 GGAAAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGCT 1380

Qy    1497 CCCCCAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAG 1556

Db    1381 CCCCCAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAG 1440

Qy    1557 AGAGAAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTG 1616

Db    1441 AGAGAAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTG 1500

Qy    1617 CACCTCAATTCTCTCTTTGGAGGAGACCAGTAG 1649

Db    1501 CACCTCAATTCTCTCTTTGGAGGAGACCAGTAG 1533


RESULT 8
AAT61085
ID   AAT61085 standard; DNA; 10163 BP.
XX
AC   AAT61085;
XX
DT   15-OCT-1997  (first entry)
XX
DE   Full-length HIV-2KR proviral DNA.
XX
KW   HIV-2; HIV-2KR; provirus; gene therapy; vector; vaccine; AIDS; ss.
XX
OS   Human immunodeficiency virus type 2.
XX
FH   Key             Location/Qualifiers
FT   provirus        1..10163
FT                   /*tag=  a
FT                   /note= "HIV-2KR"
FT   LTR             1..1070
FT                   /*tag=  b
FT                   /note= "5'LTR (Claim 14)"
FT   CDS             1071..2636
FT                   /*tag=  c
FT                   /note= "gag gene"
FT   CDS             2294..6038
FT                   /*tag=  d
FT                   /note= "pol gene"
FT   CDS             5391..6038
FT                   /*tag=  e
FT                   /note= "vif gene"
FT   CDS             6201..6518
FT                   /*tag=  f
FT                   /note= "vpr gene"
FT   CDS             6364..6650
FT                   /*tag=  g
FT                   /note= "tat gene 5' region"
FT   intron          6651..8795
FT                   /*tag=  h
FT                   /note= "intron interrupts tat and rev genes"
FT   CDS             8796..8892
FT                   /*tag=  i
FT                   /note= "tat gene 3' region"
FT   CDS             6590..6650
FT                   /*tag=  j
FT                   /note= "rev gene 5' region"
FT   CDS             8796..9265
FT                   /*tag=  k
FT                   /note= "rev gene 3' region"
FT   CDS             6660..9233
FT                   /*tag=  l
FT                   /note= "env gene"
FT   CDS             9067..9828
FT                   /*tag=  m
FT                   /note= "nef gene"
XX
PN   WO9705242-A1.
XX
PD   13-FEB-1997.
XX
PF   09-JUL-1996;   96WO-US11445.
XX
PR   07-JUN-1996;   96US-0659251.
PR   26-JUL-1995;   95US-0001441.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Kraus G,  Poeschla E,  Talbott R,  Wong-Staal F;
XX
DR   WPI; 1997-145681/13.
DR   P-PSDB; AAW13052-59.
XX
PT   Human immunodeficiency virus type II proviral clone - useful to
PT   develop prods. for study, diagnosis, prevention and treatment of HIV
PT   infections
XX
PS   Claim 1; Page 79-84; 128pp; English.
XX
CC   An isolated HIV-2 provirus (T61085), designated HIV-2KR and
CC   deposited as GenBank No. U22047, comprises a full-length HIV-2
CC   genome.  It was obtd. from the genomic DNA of Molt 4 Clone 8
CC   lymphoblastic cells infected with HIV-2PEI2, a strain isolated from
CC   the peripheral blood mononuclear cells of an HIV-2 seropositive
CC   AIDS patient.  Sequencing showed that the rev gene extends for
CC   nearly 69 amino acids past the end of most other HIV-2/SIV rev
CC   genes, and that the 5'LTR contains a deletion of 9-10 bp that is
CC   not present in other HIV-2/SIV isolates.  HIV-2KR is capable of
CC   infecting macaque peripheral blood lymphocytes in vitro and
CC   produces productive and persistent, albeit attenuated, infection in
CC   Macacca nemestrina.  Proviral constructs are useful for the
CC   generation of in vitro diagnostic reagents, gene therapy vectors
CC   and vaccines, and for the generation of HIV-2 based packaging cell
CC   lines.
XX
SQ   Sequence 10163 BP; 3456 A; 2046 C; 2540 G; 2121 T; 0 other;

  Query Match             61.4%;  Score 1174.4;  DB 18;  Length 10163;
  Best Local Similarity   77.4%;  Pred. No. 4.4e-294;
  Matches 1515;  Conservative    0;  Mismatches  381;  Indels   62;  Gaps    5;

Qy       7 AGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGC--------- 57

Db     952 AGGCGGCGTGTGGAGCGGGAG-TAAAGAGGCCTCCGGGTGAAGGTAAGTACCTACACCAA 1010

Qy      58 -AACACAAAAAAGAAATAGCTGTCTTTTATCCAGGAAGGGGTAATAAGATAGAGTGGGAG 116

Db    1011 AAACTGTAGCCAGAAAAAGGCTTGTTATCCTACCTTTAGACAGGTAGAAGATTGTGGGAG 1070

Qy     117 ATGGGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGG 176

Db    1071 ATGGGCGCGAGAAGCTCCGTCTTGAGAGGGAAAAAAGTAGATGAATTAGAAAAAATTAGG 1130

Qy     177 CTACGACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAA 236

Db    1131 TTACGGCCCGGCGGGAAGAAAAAATATAGACTAAAGCATATTGTGTGGGCAGCGAATGAA 1190

Qy     237 TTAGATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATA 296

Db    1191 TTGGGCAAATTCGGATTGGCAGAAAGCCTGTTGGAGTCAAAAGAAGGTTGCCAAAAAATT 1250

Qy     297 CTTTCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAAT 356

Db    1251 ATTACAGTTTTAGATCCATTAGTGCCAACAGGATCAGAAAATTTAAAAAGCCTTTTTAAT 1310

Qy     357 ACTGTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCA 416

Db    1311 ACTGTCTGCGTCATTTGGTGCTTGCACGCAGAAGAGAAAGTGAAAGATACTGAAGGAGCA 1370
```

```
Qy    417 AAACAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAA 476
          ||||| ||||| |||||||| |||||||  |||||||||||  |||| |  |||||||
Db   1371 AAACAAATAGTACAGAGACATCTAGTGGCAGAAACAGGAACTGCAGACAAAATGCCAAGC 1430

Qy    477 ACAAGTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATA 536
          |||||||||||| |||||||| ||||||| ||||| |||||||||| || |||||||||
Db   1431 ACAAGTAGACCAGCAGCACCACCTAGCGGGAGAGGGGGAAATTACCCCGTGCAACAAATA 1490

Qy    537 GGTGGTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTG 596
          | ||| ||||||  ||| |||| || ||||| |||  |||||| |||||||| ||
Db   1491 GCTGGCAACTATTCCCATGTGCCGTTGAGCCCCCGAACCCTAAATGCTTGGGTAAAGTTA 1550

Qy    597 ATAGAGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGT 656
            || |||||||| || || |||||||||||||||| ||||||||||| ||||||||
Db   1551 GTGGAAGAAAAGAAGTTCGGGGCAGAAGTAGTGCCAGGGTTTCAGGCACTCTCAGAAGGC 1610

Qy    657 TGCACCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATG 716
          ||||| |||||||| |||||||| ||| | |||||||||| |||||||||| ||||||
Db   1611 TGCACGCCCTATGATATTAATCAAATGCTTAATTGTGTGGGCGACCATCAAGCAGCTATG 1670

Qy    717 CAGATTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAA 776
          || || |||||||| ||||| || || || || ||| |||||||||  |||| ||||||  |
Db   1671 CAAATAATCAGAGAGATTATTAATGAAGAAGCAGCAGATTGGGATGTGCAACACCCAATA 1730

Qy    777 CCAGCTCC---ACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACA 833
          ||||  ||    | |   || || ||||| || ||   ||| || ||||| ||||| |||
Db   1731 CCAGGCCCCTTGCCAGCGGGGCAGCTTAGAGAACCAAGAGGGTCTGATATAGCAGGGACA 1790

Qy    834 ACTAGTTCAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTA 893
          || ||  ||||||| ||||| ||||||||||||||  ||| ||| || || |||||||||
Db   1791 ACAAGCACAGTAGAAGAACAGATCCAGTGGATGTTTAGAGCACAAAATCCTATACCAGTA 1850

Qy    894 GGCAACATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAAC 953
          || ||||| || |||||||||||||  | ||  |||| || || ||||| ||||| ||
Db   1851 GGGAACATCTATAGGAGATGGATCCAGATAGGACTGCAGAAGTGCGTCAGGATGTACAAT 1910

Qy    954 CCAACAAACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGAC 1013
          ||||| ||||| ||||| |||||||| || ||||| ||||| || || ||||||||||
Db   1911 CCAACCAACATCCTAGACGTAAAACAGGGACCAAAGGAGCCGTTCCAAAGCTATGTAGAT 1970

Qy    1014 AGGTTCTACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACT 1073
           || |||||||||||  |||| |||||||| |||||  |||||||||| ||||||||||
Db   1971 AGATTCTACAAAAGCCTAAGGGCAGAACAAACAGACCCAGCAGTAAAAAATTGGATGACC 2030

Qy    1074 CAAACACTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGT 1133
           |||||||||||  | || |||||  |||||||| || ||  |||| || || || |||||
Db   2031 CAAACACTGCTGGTACAGAATGCCAACCCAGACTGTAAATTAGTACTAAAAGGACTGGGG 2090

Qy    1134 GTGAATCCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAG 1193
            ||||||| ||| |||| || |||||||| || ||||| ||| |||| || || || |||
Db   2091 ATGAATCCTACCTTAGAGGAGATGCTGACCGCCTGTCAGGGAATAGGAGGACCAGGCCAG 2150

Qy    1194 AAGGCTAGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTT 1253
           || || |||||||||||||||||| | || |||||||| |||||||  || ||||| |||
Db   2151 AAAGCCAGATTAATGGCAGAAGCCTTAAAGGAGGCCCTAGCACCAGCCCCTATCCCATTT 2210

Qy    1254 GCAGCAGCCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAG 1313
           ||||||||||||||          |||| | ||||||||| |||||||||| || ||
Db   2211 GCAGCAGCCCAACAG---------AGAAGGACAATTAAGTGCTGGAATTGTGGAAAGGAT 2261

Qy    1314 GGACACTCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAA 1373
           || ||||| ||||| |||||| |||| || |||||||||| |||||||||||||| |||
Db   2262 GGGCACTCGGCAAGACAATGCCGAGCACCTAGAAGACAGGGCTGCTGGAAATGTGGCAAA 2321

Qy    1374 ATGGACCATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCA 1433
              |  ||||| ||||| || |||||||| |||||||| ||||||||||  |||| |||
Db   2322 TCAGGACATGTCATGGCAAACTGCCCAGAAAGACAGGCTGGTTTTTTAGGGATTGGCCCA 2381

Qy    1434 TGGGGAAAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACT 1493
           |||||||||||||| ||||| |||||| || | | ||| |  ||||||||||  |||||
Db   2382 TGGGGAAAGAAGCCTCGCAACTTCCCCGTGACCCGAGTCCCGCAGGGGCTGACACCAACA 2441

Qy    1494 GCTCCCCCAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAG 1553
           || |||||||  |||||||| |  || |||||| |||| ||| |||||   || | || ||
Db   2442 GCACCCCCAGCAGACCCAGCAGCAGACCTGCTAGAGAAGTACTTGCAGCAAGGGAGGAAG 2501

Qy    1554 CAGAGAGAAAAGCAGAGA----------------------------------------GAA 1574
           ||||  |||  || | |                                        |||
Db   2502 CAGAAAGAGCAGAAAATGAGACCATACAAGGAGGTGACAGAGGACTTACTGCACCTCGAA 2561

Qy    1575 AGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCACCTCAATTCTCTCTTT 1634
               ||||||  ||  |||| |||| ||||||||||||||||||||||||||||||||||
Db   2562 CAAGGAGAGACACCACACAAAGAGGCGACAGAGGATTTGCTGCACCTCAATTCTCTCTTT 2621

Qy    1635 GGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAGAAGTATTACTGGATA 1694
           |||  |||||||||||||| ||  || |||| || ||||| |||||||| ||||| || |
Db   2622 GGAAAAGACCAGTAGTCACAGCATATGTTGAGGGTCAGCCAGTAGAAGTCTTACTAGACA 2681

Qy    1695 CAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCACATTATACCCCAAAAA 1754
           |||||||||| || || || ||| |||||||||||||| ||     ||||||  |||||||
Db   2682 CAGGGGCTGACGACTCAATAGTAGCAGGAATAGAGTTGGGGAGCAATTATAGTCCAAAAA 2741

Qy    1755 TAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAAATGTAGAAATAGAAG 1814
           ||||||| |||||||| || || || || || || ||||| |||||||||||||| |||
Db   2742 TAGTAGGGGGAATAGGGGGATTCATAAACACCAAGGAATATAAAAATGTAGAAATAAAAG 2801

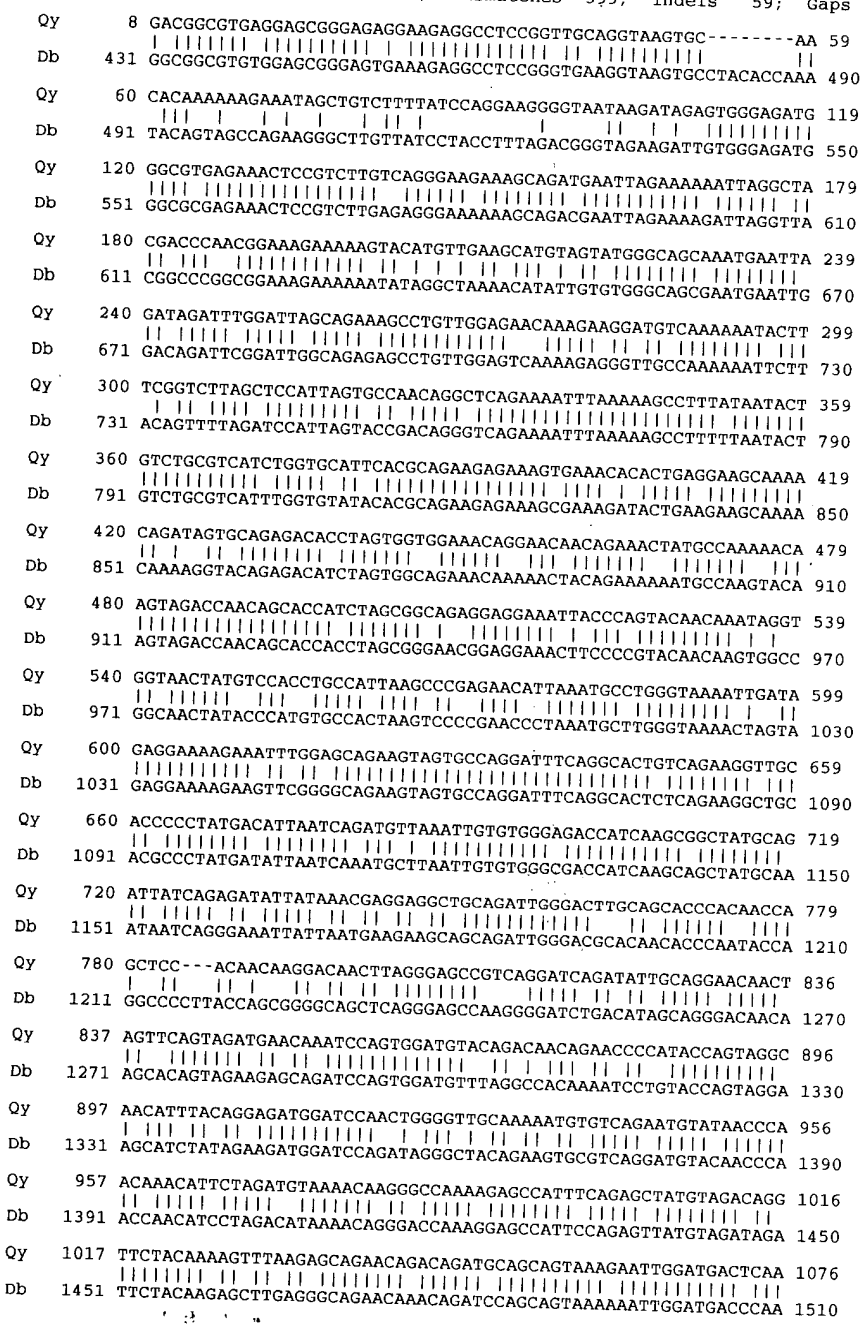
Qy    1815 TTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCCCGATTAACATTTTTG 1874
           |  ||   |||| | | |||| || || |||||||| || ||||| || ||||||||||
Db   2802 TACTAAATAAAAAGGTAAAAGCCACCATAATGACAGGTGATACCCCAATCAACATTTTTG 2861

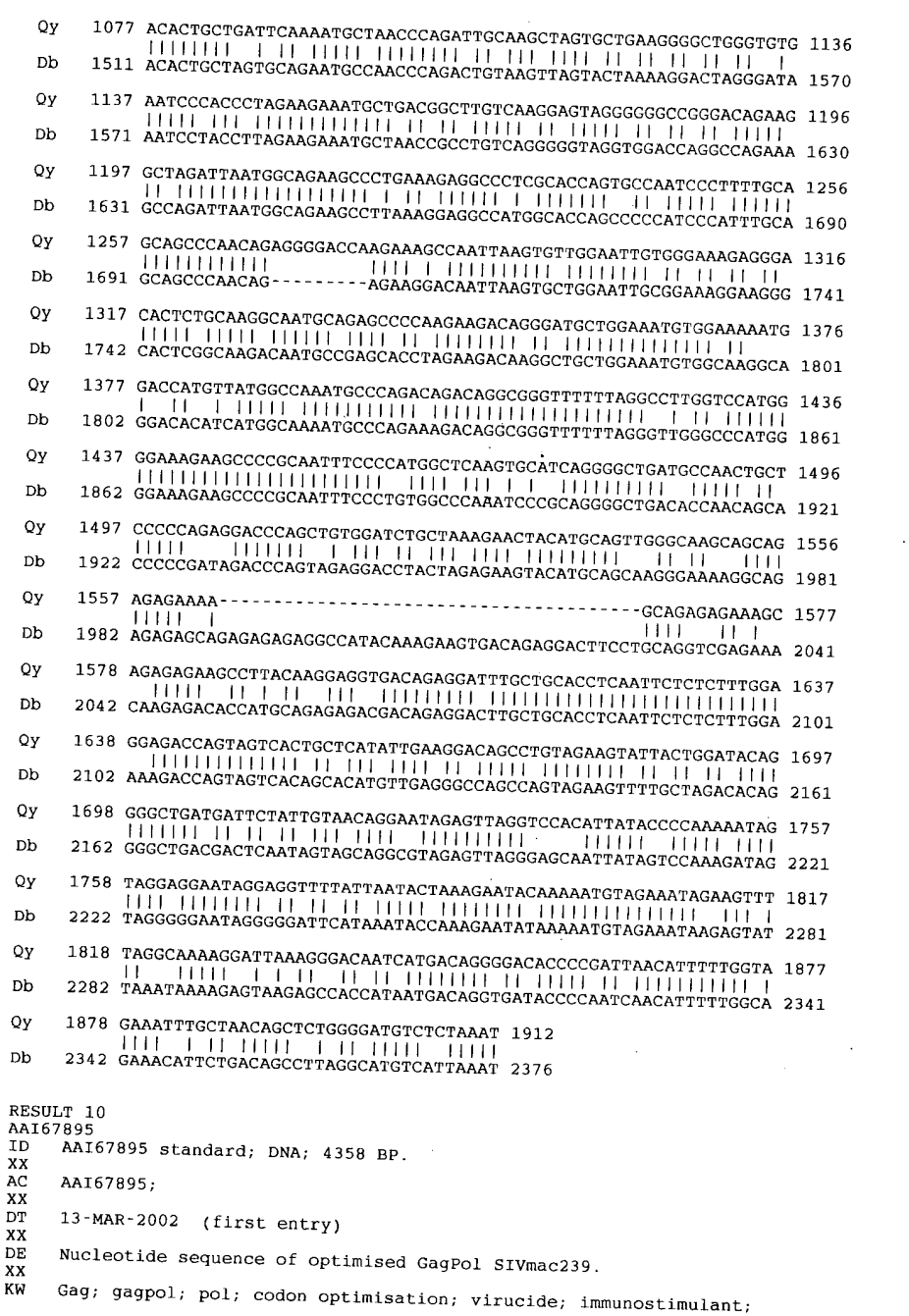
Qy    1875 GTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAAT 1912
           | |||||  | || |||||  | || |||||  |||||
Db   2862 GCAGAAACATTCTGACAGCCTTAGGCATGTCATTAAAT 2899
```

RESULT 9
AAN91774
ID AAN91774 standard; DNA; 9822 BP.
XX
AC AAN91774;
XX
DT 16-MAR-1990 (first entry)
XX
DE Entire HIV-2/ST provirus DNA.
XX
KW Antibodies
XX
OS Human immunodeficiency virus-2.
XX
PN WO8909815-A.
XX
PD 19-OCT-1989.
XX
PF 04-APR-1989; 89WO-US01379.
XX
PR 04-APR-1988; 88US-0177562.
XX
PA (RESE) RESEARCH CORPORATION TECHNOLOGIES INC.
XX
PI Shaw GM, Hahn BH, Kong LI, Lee SW;
XX
DR WPI; 1989-324223/44.
XX
PT Novel virus of the HIV-2 family - used in the prodn. of diagnostic
PT reagents, vaccines and therapies for AIDS and AIDS-related diseases
XX
PS Claim 37; page 47-53; 66pp; English.
XX
CC This is used as a source of recombinant viral components in the
CC development of diagnostic assays for HIV-2 viruses.
XX
SQ Sequence 9822 BP; 3359 A; 2023 C; 2441 G; 1999 T; 0 other;

Query Match 60.6%; Score 1158.6; DB 10; Length 9822;
Best Local Similarity 76.6%; Pred. No. 5.4e-290;

Matches 1497; Conservative 0; Mismatches 399; Indels 59; Gaps 4;





RESULT 10
AAI67895
ID AAI67895 standard; DNA; 4358 BP.
XX
AC AAI67895;
XX
DT 13-MAR-2002 (first entry)
XX
DE Nucleotide sequence of optimised GagPol SIVmac239.
XX
KW Gag; gagpol; pol; codon optimisation; virucide; immunostimulant;

KW vaccine; gene therapy; SIV; ss.
XX
OS Synthetic.
OS Simian immunodeficiency virus.
XX
PN EP1156112-A1.
XX
PD 21-NOV-2001.
XX
PF 18-MAY-2000; 2000EP-0110623.
XX
PR 18-MAY-2000; 2000EP-0110623.
XX
PA (GENE-) GENEART GMBH.
XX
PI Wagner R, Graf M, Deml L, Bieler K;
XX
DR WPI; 2002-043171/06.
XX
PT Synthetic gag and gagpol genes optimized for high level expression via
PT codon optimization, for efficient generation of vector particle useful
PT in gene delivery and for immunizing against diseases of retroviruses -
XX
PS Claim 3; Page 18-20; 40pp; English.
XX
CC The invention relates to synthetic gag and gagpol genes optimized for
CC high level expression via codon optimization. The optimised genes can be
CC used for treatment and prophylaxis of diseases caused by retroviruses as
CC well as efficient generation of vector particles (as packaging construct
CC for gene delivery). The optimized genes may also be used in the
CC construction of GagPol based vaccines modified for improved safety. When
CC used as packaging constructs,the optimised genes exhibit safe and
CC efficient gene delivery and expression of Gag of GagPol genes; and
CC allows expression of high yields of native proteins in higher eukaryotes
CC or insect cells, and supporting the generation of transduction competent
CC particles for vaccination and gene therapy purposes, unlike the complex
CC regulatory mechanisms of late gag and gappol expression found in the
CC prior art. The present sequence represents the nucleotide sequence of
CC optimised GagPol SIVmac239.
XX
SQ Sequence 4358 BP; 1059 A; 1304 C; 1422 G; 573 T; 0 other;

Query Match 60.0%; Score 1147.4; DB 24; Length 4358;
Best Local Similarity 77.4%; Pred. No. 3.1e-287;
Matches 1391; Conservative 0; Mismatches 406; Indels 0; Gaps 0;

```
Qy    117 ATGGGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGG 176
Db      1 ATGGGCGTGAGGAACAGCGTGCTGAGCGGCAAGAAGGCCGACGAGCTGGAGAAGATCAGG 60

Qy    177 CTACGACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAA 236
Db     61 CTGAGGCCCAACGGCAAGAAGAAGTATATGCTGAAGCACGTGGTGTGGGCCGCCAACGAG 120

Qy    237 TTAGATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATA 296
Db    121 CTGGACAGGTTCGGCCTGGCCGAGAGCCTGCTGGAGAACAAGGAGGGCTGCCAGAAGATC 180

Qy    297 CTTTCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAAT 356
Db    181 CTGAGCGTGCTGGCCCCCCTGGTGCCCACCGGCAGCGAGAACCTGAAGAGCCTGTACAAC 240

Qy    357 ACTGTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCA 416
Db    241 ACCGTGTGCGTGATCTGGTGCATCCACGCCGAGGAGAAGGTGAAGCACACCGAGGAGGCC 300

Qy    417 AAACAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAA 476
Db    301 AAGCAGATCGTGCAGAGGCACCTGGTGGTGGAGACCGGCACCACCGAGACCATGCCCAAG 360

Qy    477 ACAAGTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATA 536
Db    361 ACCAGCAGGCCCACCGCCCCCAGCTCCGGCCGCGGCGGCAACTACCCCGTGCAGCAGATC 420

Qy    537 GGTGGTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTG 596
Db    421 GGCGGCAACTACGTGCACCTGCCCCTGAGCCCCAGGACCCTGAACGCCTGGGTGAAGCTG 480

Qy    597 ATAGAGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGT 656
Db    481 ATCGAGGAGAAGAAGTTCGGCGCCGAGGTGGTGCCCGGCTTCCAGGCCCTGAGCGAGGGC 540

Qy    657 TGCACCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATG 716
Db    541 TGCACCCCTTACGACATCAACCAGATGCTGAACTGCGTGGGCGACCACCAGGCCGCCATG 600

Qy    717 CAGATTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAA 776
Db    601 CAGATCATCAGGGACATCATCAACGAGGAGGCCGCCGACTGGGACCTGCAGCACCCTCAG 660

Qy    777 CCAGCTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACT 836
Db    661 CCCGCCCCTCAGCAGGGCCAGCTGAGGGAGCCCAGCGGCAGCGACATCGCCGGCACCACA 720

Qy    837 AGTTCAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGC 896
Db    721 AGCAGCGTGGACGAGCAGATCCAGTGGATGTACAGGCAGCAGAACCCTATCCCCGTGGGC 780

Qy    897 AACATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCA 956
Db    781 AACATCTACAGGAGGTGGATCCAGCTGGGCCTCCAGAAGTGCGTGAGGATGTACAACCCC 840

Qy    957 ACAAACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGG 1016
Db    841 ACAAACATCCTGGACGTGAAGCAGGGACCAAAGGAGCCCTTCCAGTCATATGTGGACAGG 900

Qy   1017 TTCTACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAA 1076
Db    901 TTCTACAAGAGCCTGAGGGCCGAGCAGACCGACGCCGCCGTGAAGAACTGGATGACCCAG 960

Qy   1077 ACACTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTG 1136
Db    961 ACCCTGCTGATCCAGAACGCCAACCCCGACTGCAAGCTGGTGCTGAAGGGCCTGGGCGTG 1020

Qy   1137 AATCCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAG 1196
Db   1021 AACCCCACCCTGGAGGAGATGCTGACCGCCTGCCAGGGCGTGGGCGGCCCCGGCCAGAAG 1080

Qy   1197 GCTAGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCA 1256
Db   1081 GCTAGGCTGATGGCCGAGGCTCTGAAGGAGGCCCTGGCCCCCGTGCCCATCCCCTTCGCC 1140

Qy   1257 GCAGCCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGA 1316
Db   1141 GCCGCCCAGCAGAGGGGACCCAGGAAGCCCATCAAGTGCTGGAACTGCGGCAAGGAGGGC 1200

Qy   1317 CACTCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATG 1376
Db   1201 CACAGCGCCAGGCAGTGCAGGGCCCCCAGGAGGCAGGGCTGCTGGAAGTGCGGCAAGATG 1260

Qy   1377 GACCATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGG 1436
Db   1261 GACCACGTGATGGCCAAGTGCCCCGACAGGCAGGCCGGTTTTTTAGGCCTTGGTCCATGG 1320

Qy   1437 GGAAAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGCT 1496
Db   1321 GGAAAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGCT 1380

Qy   1497 CCCCCAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAG 1556
Db   1381 CCCCCAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAG 1440

Qy   1557 AGAGAAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTG 1616
Db   1441 AGAGAAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTG 1500
```

```
Qy   1617 CACCTCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTG 1676
          ||||||||||||||||||||||||||||||||||||||| || || || || || || ||||| |
Db   1501 CACCTCAATTCTCTCTTTGGAGGAGACCAGTAGTGACCGCCCACATCGAGGGCCAGCCCG 1560

Qy   1677 TAGAAGTATTACTGGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTC 1736
          | || ||  | ||||| || || || || ||    || || || || || ||| | || |
Db   1561 TGGAGGTGCTGCTGGACACCGGCGCCGACGACAGCATCGTGACCGGCATCGAGCTGGGAC 1620

Qy   1737 CACATTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACA 1796
          | || || ||||| || || || || || || || || || || || || || || ||||
Db   1621 CCCACTACACCCCCAAGATCGTGGGCGGCATCGGCGGCTTCATCAACACAAAGGAGTACA 1680

Qy   1797 AAAATGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACA 1856
          | || || || || || ||  | ||||| ||||| || || || |||||||| || ||||
Db   1681 AGAACGTGGAGATCGAGGTGCTGGGCAAGAGGATCAAGGGCACCATCATGACCGGCGACA 1740

Qy   1857 CCCCGATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAATT 1913
          |||| || ||||| || || || ||  |||| || || ||||| |||   || || |
Db   1741 CCCCCATCAACATCTTCGGCAGGAACCTGCTGACCGCCCTGGGCATGAGCCTGAACT 1797


RESULT 11
AAQ20616
ID   AAQ20616 standard; DNA; 9672 BP.
XX
AC   AAQ20616;
XX
DT   19-MAY-1992  (first entry)
XX
DE   ROD HIV-2 isolate complete genome.
XX
KW   Human immunodeficiency virus; AIDS; ss.
XX
OS   Human immunodeficiency virus-2 ROD isolate.
XX
FH   Key             Location/Qualifiers
FT   CDS             546..2111
FT                   /*tag=  a
FT                   /note= "GAG gene"
FT   Misc_feature    546..950
FT                   /*tag=  b
FT                   /note= "GAG p 16"
FT   Misc_feature    951..1640
FT                   /*tag=  c
FT                   /note= "GAG p 26"
FT   Misc_feature    1701..2111
FT                   /*tag=  d
FT                   /note= "GAG p 12"
FT   Misc_feature    1829..4936
FT                   /*tag=  e
FT                   /note= "polymerase"
FT   Misc_feature    4869..5513
FT                   /*tag=  f
FT                   /note= "Q protein"
FT   Misc_feature    5682..5996
FT                   /*tag=  g
FT                   /note= "R protein"
FT   Misc_feature    5344..5679
FT                   /*tag=  h
FT                   /note= "X protein"
FT   Misc_feature    5682..5996
FT                   /*tag=  i
FT                   /note= "Y protein"
FT   Misc_feature    6147..8720
FT                   /*tag=  j
FT                   /note= "Env protein"
FT   Misc_feature    8557..9324
FT                   /*tag=  k
FT                   /note= "F protein"
FT   Exon            5845..6140
FT                   /*tag=  l
FT                   /note= "TAT gene exon 1"
FT   Exon            8307..8400
FT                   /*tag=  m
FT                   /note= "TAT gene exon 2"
FT   Exon            6071..6140
FT                   /*tag=  n
FT                   /note= "ART gene exon 1"
FT   Exon            8307..8536
FT                   /*tag=  o
FT                   /note= "ART gene exon 2"
FT   LTR             1..173
FT                   /*tag=  p
FT                   /note= "LTR:R 1"
FT   LTR             9498..9671
FT                   /*tag=  q
FT                   /note= "LTR:R 2"
FT   Misc_feature    174..299
FT                   /*tag=  r
FT                   /note= "U5"
FT   Misc_feature    8942..9497
FT                   /*tag=  s
FT                   /note= "U3"
XX
PN   US5079342-A.
XX
PD   07-JAN-1992.
XX
PF   11-FEB-1987;   87US-0U13477.
XX
PR   11-FEB-1987;   87US-0013477.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Alizon M,  Montagnier L,  Geutard D,  Clavel F,  Sonigo P,  Guyader M;
XX
DR   WPI; 1992-041067/05.
DR   P-PSDB; AAR20596, AAR20597, AAR20598, AAR20599, AAR20600, AAR20601,
DR   AAR20602, AAR20603, AAR20604, AAR20605, AAR20606, AAR20607.
XX
PT   Peptide(s) corresp. to HIV-2 amino acid sequences - used in
PT   diagnosis in vaccines and in prodn. of antibodies for diagnosis
XX
PS   Disclosure; Page 12; 30pp; English.
XX
CC   The sequence is that of the complete genomic sequence of the ROD HIV-2
CC   isolate. Its genes and their major exprèssion products are indicated
CC   in the features. It was isolated from a CEM cell line infected with
CC   the ROD isolate.
XX
SQ   Sequence 9672 BP; 3314 A; 1981 C; 2394 G; 1983 T; 0 other;

  Query Match             59.9%;  Score 1146.6;  DB 13;  Length 9672;
  Best Local Similarity   76.5%;  Pred. No. 7e-287;
  Matches 1498;  Conservative    0;  Mismatches  399;  Indels   62;  Gaps    5;

Qy      7 AGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGCAACACAAAA 66
            || | ||||| ||||||||||  |||||||||||||| || |||||||| |       ||
Db    428 AGGCAGCGTGTGGAGCGGGAGGAGAAGAGGCCTCCGGGTGAAGGTAAGTACCTACACCAA 487

Qy     67 AAGAAATAGCTG--------TCTTTTATCCAGGAAGGGGTAATAAGATAGAGTGGGAGAT 118
          ||     |||| |        |  | |        |     ||  | |  |||||||||
Db    488 AAACTGTAGCCGAAAGGGCTTGCTATCCTACCTTTAGACAGGTAGAAGATTGTGGGAGAT 547

Qy    119 GGGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCT 178
          ||||| ||||||||||||||||  |||||| ||||||||||||||||||| ||| ||| |
Db    548 GGGCGCGAGAAACTCCGTCTTGAGAGGGAAAAAAGCAGATGAATTAGAAAGAATCAGGTT 607

Qy    179 ACGACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATT 238
          ||| |||  |||||||||||||||| | | || ||| | || ||||||| ||| ||||
Db    608 ACGGCCCGGCGGAAAGAAAAAGTACAGGCTAAAACATATTGTGTGGGCAGCGAATAAATT 667

Qy    239 AGATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACT 298
           || ||||| |||||||||| ||||||||||   ||||| || |||||||||| ||
Db    668 GGACAGATTCGGATTAGCAGAGAGCCTGTTGGAGTCAAAAGAGGGTTGTCAAAAAATTCT 727
```

```
Qy    299 TTCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATAC 358
Db    728 TACAGTTTTAGATCCAATGGTACCGACAGGTTCAGAAAATTTAAAAAGTCTTTTTAATAC 787

Qy    359 TGTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAA 418
Db    788 TGTCTGCGTCATTTGGTGCATACACGCAGAAGAGAAAGTGAAAGATACTGAAGGAGCAAA 847

Qy    419 ACAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAAAC 478
Db    848 ACAAATAGTGCGGAGACATCTAGTGGCAGAAACAGGAACTGCAGAGAAAATGCCAAGCAC 907

Qy    479 AAGTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGG 538
Db    908 AAGTAGACCAACAGCACCATCTAGCGAGAAGGGAGGAAATTACCCAGTGCAACATGTAGG 967

Qy    539 TGGTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGAT 598
Db    968 CGGCAACTACACCCATATACCGCTGAGTCCCCGAACCCTAAATGCCTGGGTAAAATTAGT 1027

Qy    599 AGAGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGTTG 658
Db   1028 AGAGGAAAAAAAGTTCGGGGCAGAAGTAGTGCCAGGATTTCAGGCACTCTCAGAAGGCTG 1087

Qy    659 CACCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCA 718
Db   1088 CACGCCCTATGATATCAACCAAATGCTTAATTGTGTGGGCGACCATCAAGCAGCCATGCA 1147

Qy    719 GATTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACC 778
Db   1148 GATAATCAGGGAGATTATCAATGAGGAAGCAGCAGAATGGGATGTGCAACATCCAATACC 1207

Qy    779 AGCTCC---ACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAAC 835
Db   1208 AGGCCCCTTACCAGCGGGGCAGCTTAGAGAGCCAAGGGGATCTGACATAGCAGGGACAAC 1267

Qy    836 TAGTTCAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGG 895
Db   1268 AAGCACAGTAGAAGAACAGATCCAGTGGATGTTTAGGCCACAAAATCCTGTACCAGTAGG 1327

Qy    896 CAACATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCC 955
Db   1328 AAACATCTATAGAAGATGGATCCAGATAGGATTGCAGAAGTGTGTCAGGATGTACAACCC 1387

Qy    956 AACAAACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAG 1015
Db   1388 GACCAACATCCTAGACATAAAACAGGGACCAAAGGAGCCGTTCCAAAGCTATGTAGATAG 1447

Qy   1016 GTTCTACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCA 1075
Db   1448 ATTCTACAAAAGCTTGAGGGCAGAACAAACAGATCCAGCAGTGAAGAATTGGATGACCCA 1507

Qy   1076 AACACTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGT 1135
Db   1508 AACACTGCTAGTACAAAATGCCAACCCAGACTGTAAATTAGTGCTAAAAGGACTAGGGAT 1567

Qy   1136 GAATCCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAA 1195
Db   1568 GAACCCTACCTTAGAAGAGATGCTGACCGCCTGTCAGGGGGTAGGTGGGCCAGGCCAGAA 1627

Qy   1196 GGCTAGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGC 1255
Db   1628 AGCTAGATTAATGGCAGAGGCCCTGAAAGAGGTCATAGGACCTGCCCCTATCCCATTCGC 1687

Qy   1256 AGCAGCCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGG 1315
Db   1688 AGCAGCCCAGCAG---------AGAAAGGCATTTAAATGCTGGAACTGTGGAAAGAAAGG 1738

Qy   1316 ACACTCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAAT 1375
Db   1739 GCACTCGGCAAGACAATGCCGAGCACCTAGAAGGCAGGGCTGCTGGAAGTGTGGTAAGCC 1798

Qy   1376 GGACCATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATG 1435

Db   1799 AGGACACATCATGACAAACTGCCCAGATAGACAGGCAGGTTTTTTAGGACTGGGCCCTTG 1858

Qy   1436 GGGAAAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGC 1495
Db   1859 GGGAAAGAAGCCCCGCAACTTCCCCGTGGCCCAAGTTCCGCAGGGGCTGACACCAACAGC 1918

Qy   1496 TCCCCCAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCA 1555
Db   1919 ACCCCCAGTGGATCCAGCAGTGGATCTACTGGAGAAATATATGCAGCAAGGGAAAAGACA 1978

Qy   1556 GAGAGAAAA----------------------------------------GCAGAGAGAAAG 1576
Db   1979 GAGAGAGCAGAGAGAGAGACCATACAAGGAAGTGACAGAGGACTTACTGCACCTCGAGCA 2038

Qy   1577 CAGAGAGAAGCCTTACAAGGAG---GTGACAGAGGATTTGCTGCACCTCAATTCTCTCTT 1633
Db   2039 GGGGGAGACACCATACAGGGAGCCACCAACAGAGGACTTGCTGCACCTCAATTCTCTCTT 2098

Qy   1634 TGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAGAAGTATTACTGGAT 1693
Db   2099 TGGAAAAGACCAGTAGTCACAGCATACATTGAGGGTCAGCCAGTAGAAGTCTTGTTAGAC 2158

Qy   1694 ACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCACATTATACCCCAAAA 1753
Db   2159 ACAGGGGCTGACGACTCAATAGTAGCAGGAATAGAGTTAGGGAACAATTATAGCCCAAAA 2218

Qy   1754 ATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAAATGTAGAAATAGAA 1813
Db   2219 ATAGTAGGGGGAATAGGGGGATTCATAAATACCAAGGAATATAAAAATGTAGAAATAGAA 2278

Qy   1814 GTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCCCGATTAACATTTTT 1873
Db   2279 GTTCTAAATAAAAAGGTACGGGCCACCATAATGACAGGCGACACCCCAATCAACATTTTT 2338

Qy   1874 GGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAAT 1912
Db   2339 GGCAGAAATATTCTGACAGCCTTAGGCATGTCATTAAAT 2377
```

RESULT 12
AAZ89628
ID AAZ89628 standard; DNA; 9630 BP.
XX
AC AAZ89628;
XX
DT 11-JUL-2000 (first entry)
XX
DE HIV-2 ROD DNA.
XX
KW Diagnosis; probe; lymphocyte; virus; immunodiagnostic; infection;
KW antiviral; ss.
XX
OS Human immunodeficiency virus type 2.
XX
PN US6054565-A.
XX
PD 25-APR-2000.
XX
PF 28-APR-1994; 94US-0234875.
XX
PR 11-FEB-1987; 87US-0013477.
PR 03-SEP-1991; 91US-0752368.
PR 20-DEC-1991; 91US-0810908.
PR 03-MAR-1986; 86US-0835228.
PR 06-OCT-1986; 86US-0916080.
PR 21-NOV-1986; 86US-0933184.
PR 16-JAN-1987; 87US-0003764.
XX
PA (INSP) INST PASTEUR.
XX
PI Montagnier L, Clavel F, Guyader M, Geutard D, Sonigo P, Alizon M;
XX

DR WPI; 2000-328365/28.
XX
PT Novel cloned nucleotide sequences homologous or identical to the
PT portion of genomic RNA of HIV-2 viruses useful as probes and in
PT diagnostic tests to diagnose HIV-2 infection -
XX
PS Claim 1; Column 10-23; 33pp; English.
XX
CC This invention describes a novel cloned nucleic acid (I) of a human
CC immunodeficiency virus type 2 (HIV-2). (I) is capable of being used
CC as probes in diagnostic method to obtain the immunological reagents
CC necessary to diagnose an HIV-2 infection. These sequences may be used
CC as probes in hybridization reactions with the genetic material of
CC infected patients to indicate whether the RNA of the HIV-2 virus is
CC present in these patient's lymphocytes or whether an analogous DNA is
CC present. The genetic sequence of the HIV-2 virus may be used to create
CC the polypeptides encoded by these sequences. Specifically, these
CC polypeptides may be created by expression of the cDNA obtained from
CC bacterial, yeast or animal cells. These polypeptides may be used in
CC diagnostic tests such as immunofluorescence assays, radioimmunoassays
CC (RIA) and Western Blot tests. Monoclonal antibodies to these polypeptides
CC of fragments may be created and used in immunodiagnostic tests. The
CC polypeptides of the present invention may also be used as immunogenic
CC reagents to induce protection against infection by HIV-2 viruses. The
CC polypeptides produced by recombinant-DNA techniques would function as
CC vaccine agents. The polypeptides may be used on competitive assays to
CC test the ability of various antiviral agents to determined their ability
CC to prevent the virus from fixing on its target. This sequence encodes
CC the HIV-2 ROD genome which is described in the method of the invention.
XX
SQ Sequence 9630 BP; 3310 A; 1959 C; 2373 G; 1976 T; 12 other;

Query Match 59.9%; Score 1145; DB 21; Length 9630;
Best Local Similarity 76.4%; Pred. No. 1.8e-286;
Matches 1497; Conservative 0; Mismatches 400; Indels 62; Gaps 5;

```
Qy      7 AGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGCAACACAAAA 66
          || | ||||| |||||||||  ||||||||||||| || ||||||||  |        ||
Db    427 AGGCAGCGTGTGGAGCGGGAGGAGAAGAGGCCTCCGGGTGAAGGTAAGTACCTACACCAA 486

Qy     67 AAGAAATAGCTG--------TCTTTTATCCAGGAAGGGGTAATAAGATAGAGTGGGAGAT 118
          ||    |||| |        | || |      |    ||  | |  |||||||||
Db    487 AAACTGTAGCCGAAAGGGCTTGTTATCCTACCTTTAGACAGGTAGAAGATTGTGGGAGAT 546

Qy    119 GGGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCT 178
          |||||  ||||||||||||||||  |||||| |||||||||||||||||||| ||| ||| |
Db    547 GGGCGCGAGAAACTCCGTCTTGAGAGGGAAAAAAGCAGATGAATTAGAAAGAATCAGGTT 606

Qy    179 ACGACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATT 238
          |||  |||  |||||||||||||||||| | | || ||| | || |||||||| ||| ||||
Db    607 ACGGCCCGGCGGAAAGAAAAAGTACAGGCTAAAACATATTGTGTGGGCAGCGAATAAATT 666

Qy    239 AGATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACT 298
          || ||||| |||||||||| ||||||||||    ||||| || ||||||||||  ||
Db    667 GGACAGATTCGGATTAGCAGAGAGCCTGTTGGACTCAAAAGAGGGTTGTCAAAAAATTCT 726

Qy    299 TTCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATAC 358
          | | || |||| |||| | || || ||||| |||||||||||||||| |||| ||||||
Db    727 TACAGTTTTAGATCCAATGGTACCGACAGGTTCAGAAAATTTAAAAAGTCTTTTTAATAC 786

Qy    359 TGTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAA 418
          |||||||||||| |||||||| |||||||||||||||||||| | ||||| | ||||||
Db    787 TGTCTGCGTCATTTGGTGCATACACGCAGAAGAGAAAGTGAAAGATACTGAAGGAGCAAA 846

Qy    419 ACAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAAAC 478
          ||| ||||||| |||||| |||||||  ||||||||||  |||| |  ||||||||  ||
Db    847 ACAAATAGTGCGGAGACATCTAGTGGCAGAAACAGGAACTGCAGAGAAAATGCCAAGCAC 906

Qy    479 AAGTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGG 538
          |||||||||||||||||||||||  |  ||||||||||||||||  |||||  ||||
Db    907 AAGTAGACCAACAGCACCATCTAGCGAGAAGGGAGGAAATTACCCAGTGCAACATGTAGG 966

Qy    539 TGGTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGAT 598
```

```
Db    967 CGGCAACTACACCCATATACCGCTGAGTCCCCGAACCCATAATGCCTGGGTAAAATTAGT 1026

Qy    599 AGAGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGTTG 658
          |||||||||  || || || ||||||||||||||||||||||||||||||||||||||||
Db   1027 AGAGGAAAAAAAGTTCGGGGCAGAAGTAGTGCCAGGATTTCAGGCACTCTCAGAAGGCTG 1086

Qy    659 CACCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCA 718
          ||| |||||||| || || || ||| | |||||||||| |||||||||| || |||||
Db   1087 CACGCCCTATGATATCAACCAAATGCTTAATTGTGTGGGCGACCATCAAGCAGCCATGCA 1146

Qy    719 GATTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACC 778
          ||| ||||| || ||||| || ||||| || |||||  |||| || |||  |||
Db   1147 GATAATCAGGGAGATTATCAATGAGGAAGCAGCAGAATGGGATGTGCAACATCCAATACC 1206

Qy    779 AGCTCC---ACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAAC 835
          ||  ||   || |  || || ||||| |||||    ||||| || || ||||| |||||
Db   1207 AGGCCCCTTACCAGCGGGGCAGCTTAGAGAGCCAAGGGGATCTGACATAGCAGGGACAAC 1266

Qy    836 TAGTTCAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGG 895
          ||  ||||||| |||| |||||||||||||  || | ||| || ||  ||||||||||
Db   1267 AAGCACAGTAGAAGAACAGATCCAGTGGATGTTTAGGCCACAAAATCCTGTACCAGTAGG 1326

Qy    896 CAACATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCC 955
          |||||  || || ||||||||||  | || ||||| || |||||||| |||| |||||
Db   1327 AAACATCTATAGAAGATGGATCCAGATAGGATTGCAGAAGTGTGTCAGGATGTACAACCC 1386

Qy    956 AACAAACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAG 1015
          || ||||| |||||  ||||||| || ||||| ||||| || || ||||||||||| ||
Db   1387 GACCAACATCCTAGACATAAAACAGGGACCAAAGGAGCCGTTCCAAAGCTATGTAGATAG 1446

Qy   1016 GTTCTACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCA 1075
          ||||||||||| || || |||||||| ||||| ||||||| |||||||||||||||| ||
Db   1447 ATTCTACAAAAGCTTGAGGGCAGAACAAACAGATCCAGCAGTGAAGAATTGGATGACCCA 1506

Qy   1076 AACACTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGT 1135
          |||||||||  | |||||||| |||||||| || ||  ||||||| || || || ||  |
Db   1507 AACACTGCTAGTACAAAATGCCAACCCAGACTGTAAATTAGTGCTAAAAGGACTAGGGAT 1566

Qy   1136 GAATCCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAA 1195
          ||| || ||| ||||||| |||||||| || ||||| || ||||| ||||| || |||||
Db   1567 GAACCCTACCTTAGAAGAGATGCTGACCGCCTGTCAGGGGGTAGGTGGGCCAGGCCAGAA 1626

Qy   1196 GGCTAGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGC 1255
          |||||||||||||||| |||||||||||| | | ||| |  || ||||| || ||
Db   1627 AGCTAGATTAATGGCAGAGGCCCTGAAAGAGGTCATAGGACCTGCCCCTATCCCATTCGC 1686

Qy   1256 AGCAGCCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGG 1315
          |||||||| |||          |||||| || |||| || ||||| ||||| || || ||
Db   1687 AGCAGCCCAGCAG---------AGAAAGGCATTTAAATGCTGGAACTGTGGAAAGGAAGG 1737

Qy   1316 ACACTCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAAT 1375
          ||||| ||||| |||||| |||| || ||||| ||||| |||||||| |||| ||
Db   1738 GCACTCGGCAAGACAATGCCGAGCACCTAGAAGGCAGGGCTGCTGGAAGTGTGGTAAGCC 1797

Qy   1376 GGACCATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATG 1435
            | ||  | ||| | || |||||||| |||||||| |||||||||| || || || ||
Db   1798 AGGACACATCATGACAAAGTGCCCAGATAGACAGGCAGGTTTTTTAGGACTGGGCCCTTG 1857

Qy   1436 GGGAAAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGC 1495
          |||||||||||||||||| ||||| |||| ||||| |  |||||||||  |||| ||
Db   1858 GGGAAAGAAGCCCCGCAACTTCCCCGTGGCCCAAGTTCCGCAGGGGCTGACACCAACAGC 1917

Qy   1496 TCCCCCAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCA 1555
          |||||||  ||| ||||| |||||||| ||  |||| || ||||||   || ||   ||
Db   1918 ACCCCCAGTGGATCCAGCAGTGGATCTACTGGAGAAATATATGCAGCAAGGGAAAAGACA 1977

Qy   1556 GAGAGAAAA----------------------------------------GCAGAGAGAAAG 1576
          ||||||  |                                        |||    ||
Db   1978 GAGAGAGCAGAGAGAGAGACCATACAAGGAAGTGACAGAGGACTTACTGCACCTCGAGCA 2037

Qy   1577 CAGAGAGAAGCCTTACAAGGAG---GTGACAGAGGATTTGCTGCACCTCAATTCTCTCTT 1633
            | ||||  || |||| ||||      |||||||| ||||||||||||||||||||||||
```

```
Db   2038 GGGGGAGACACCATACAGGGAGCCACCAACAGAGGACTTGCTGCACCTCAATTCTCTCTT 2097

Qy   1634 TGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAGAAGTATTACTGGAT 1693
          ||||  |||||||||||||| ||  | ||||| || ||||| |||||||| ||  | ||
Db   2098 TGGAAAAGACCAGTAGTCACAGCATACATTGAGGGTCAGCCAGTAGAAGTCTTGTTAGAC 2157

Qy   1694 ACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCACATTATACCCCAAAA 1753
          |||||||||| || || || ||| ||||||||||||||||     |||||| |||||||
Db   2158 ACAGGGGCTGACGACTCAATAGTAGCAGGAATAGAGTTAGGGAACAATTATAGCCCAAAA 2217

Qy   1754 ATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAAATGTAGAAATAGAA 1813
          |||||||| |||||||| || || || ||||| || |||||.||||||||||||||||||
Db   2218 ATAGTAGGGGGAATAGGGGGATTCATAAATACCAAGGAATATAAAAATGTAGAAATAGAA 2277

Qy   1814 GTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCCCGATTAACATTTTT 1873
          ||| ||   |||| | |    |  || || |||||||| |||||||| || |||||||||
Db   2278 GTTCTAAATAAAAAGGTACGGGCCACCATAATGACAGGCGACACCCCAATCAACATTTTT 2337

Qy   1874 GGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAAT 1912
          || |||||| | || |||||  | || |||||  |||||
Db   2338 GGCAGAAATATTCTGACAGCCTTAGGCATGTCATTAAAT 2376


RESULT 13
AAN80859
ID   AAN80859 standard; cDNA; 9643 BP.
XX
AC   AAN80859;
XX
DT   15-NOV-1990  (first entry)
XX
DE   Sequence of entire HIV-2 ROD genome.
XX
KW   LAV-II ROD; ss.
XX
OS   Human immunodeficiency virus ROD.
XX
FH   Key             Location/Qualifiers
FT   CDS             545..2110
FT                   /*tag= a
FT                   /note="AAP80801"
FT   CDS             1828..4908
FT                   /*tag= b
FT                   /note="AAP81773"
FT   CDS             4841..5485
FT                   /*tag= c
FT                   /note="AAP81774"
FT   CDS             5316..5654
FT                   /*tag= d
FT                   /note="AAP81775"
FT   CDS             5654..5968
FT                   /*tag= e
FT                   /note="AAP81776"
FT   CDS             5817..6113
FT                   /*tag= f
FT                   /note="AAP81777"
FT   CDS             6043..6153
FT                   /*tag= g
FT                   /note="AAP81778"
FT   CDS             6119..8692
FT                   /*tag= h
FT                   /note="AAP81779"
FT   CDS             8220..8372
FT                   /*tag= i
FT                   /note="AAP81780"
FT   CDS             8242..8508
FT                   /*tag= j
FT                   /note="AAP81781"
FT   CDS             8529..9299
FT                   /*tag= k
FT                   /note="AAP81782"
XX
PN   WO8805440-A.
XX
PD   28-JUL-1988.
XX
PF   15-JAN-1988;   88WO-FR00025.
XX
PR   15-APR-1987;   87FR-0005398.
XX
PA   (INSP ) INST PASTEUR.
PA   (ALIZ/) ALIZON M.
XX
PI   Alizon M,  Montagnier L,  Guetard D,  Clavfl F,  Sonigo P,  Guyader M;
PI   Tiollais P,  Chakrabarti L,  Desrosiers R;
XX
DR   WPI; 1988-220290/31.
DR   P-PSDB; AAP80801, AAP81773, AAP81774, AAP81775, AAP81776, AAP81777,
DR   AAP81778, AAP81779, AAP81780, AAP81781, AAP81782.
XX
PT   New peptide(s) with immunological properties of HIV-2 envelope protein -
PT   having the structure of simian immune deficiency virus proteins,
PT   useful in diagnosis and of vaccine components
XX
PS   Disclosure; Fig 1A; 86pp; French.
XX
CC   The SQ in AAN80859 was deposited on 21/2/86 at the CNCM under number
CC   I-522, reference name LAV-II ROD. It is the cDNA to HIV-2 ROD genomic
CC   RNA. SQ was compared with the SQ of the genome of SIV (Mac) (AAN80860) to
CC   identify common regions.
XX
SQ   Sequence 9643 BP; 3296 A; 1969 C; 2399 G; 1979 T; 0 other;

  Query Match           59.8%;  Score 1143.4;  DB 9;  Length 9643;
  Best Local Similarity   76.4%;  Pred. No. 4.7e-286;
  Matches 1496;  Conservative    0;  Mismatches  401;  Indels   62;  Gaps    5;

Qy      7 AGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGCAACACAAAA 66
          || | ||||| ||||||||||  |||||||||||||| || |||||||| |      ||
Db    427 AGGCAGCGTGTGGAGCGGGAGGAGAAGAGGCCTCCGGGTGAAGGTAAGTACCTACACCAA 486

Qy     67 AAGAAATAGCTG--------TCTTTTATCCAGGAAGGGGTAATAAGATAGAGTGGGAGAT 118
          ||    |||| |        |  | |          ||  | |  |||||||||
Db    487 AAACTGTAGCCGAAAGGGCTTGCTATCCTACCTTTACACAGGTAGAAGATTGTGGGAGAT 546

Qy    119 GGGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCT 178
          ||||| |||||||||||||||  |||||| |||||||||||||||||| ||| ||| |
Db    547 GGGCGCGAGAAACTCCGTCTTGAGAGGGAAAAAAGCAGATGAATTAGAAAGAATCAGGTT 606

Qy    179 ACGACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATT 238
          ||| |||  ||||||||||||||||| | | || ||| | || |||||||| ||| ||||
Db    607 ACGGCCCGGCGGAAAGAAAAAGTACAGGCTAAAACATATTGTGTGGGCAGCGAATAAATT 666

Qy    239 AGATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACT 298
           || ||||| |||||||||||  ||||||||||||   ||||| || | |||||||| ||
Db    667 GGACAGATTGGGATTAGCAGAGAGCCTGTTGGAGTCAAAAGAGGGTTCTCAAAAAATTCT 726

Qy    299 TTCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATAC 358
          | | || |||| |||| | || || ||||| ||||||||||||||||| |||| ||||||
Db    727 TACAGTTTTAGATCCAATGGTACCGACAGGTTCAGAAAATTTAAAAAGTCTTTTTAATAC 786

Qy    359 TGTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAA 418
          |||||||||||| |||||||| |||||||||||||||||||| | ||||| | ||||||
Db    787 TGTCTGCGTCATTTGGTGCATACACGCAGAAGAGAAAGTGAAAGATACTGAAGGAGCAAA 846

Qy    419 ACAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAAAC 478
          ||| ||||||| |||||| |||||||  |||||||||| |||| |  ||||||| ||
Db    847 ACAAATAGTGCGGAGACATCTAGTGGCAGAAACAGGAACTGCAGAGAAAATGCCAAGCAC 906

Qy    479 AAGTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGG 538
          |||||||||||||||||||||||||  | ||||||||||||||||| ||||| ||||
Db    907 AAGTAGACCAACAGCACCATCTAGCGAGAAGGGAGGAAATTACCCAGTGCAACATGTAGG 966

Qy    539 TGGTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTGAT 598
          || |||||   |||  | ||  | || ||  ||||  ||||||||||||||||||  |
Db    967 CGGCAACTACACCCATATACCGCTGAGTCCCCGAACCCTAAATGCCTGGGTAAAATTAGT 1026
```

```
Qy     599 AGAGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGTTG 658
           ||||||||| || || || ||||||||||||||||||||||||||||||| |||||||| ||
Db    1027 AGAGGAAAAAAAGTTCGGGGCAGAAGTAGTGCCAGGATTTCAGGCACTCTCAGAAGGCTG 1086

Qy     659 CACCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATGCA 718
           ||| |||||||| || || || ||| | |||||||||| |||||||||| || |||||
Db    1087 CACGCCCTATGATATCAACCAAATGCTTAATTGTGTGGGCGACCATCAAGCAGCCATGCA 1146

Qy     719 GATTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAACC 778
           ||| ||||| || ||||| || ||||| || ||||| |||||  |||| || |||  |||
Db    1147 GATAATCAGGGAGATTATCAATGAGGAAGCAGCAGAATGGGATGTGCAACATCCAATACC 1206

Qy     779 AGCTCC---ACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAAC 835
           ||  ||   || |   || || ||||| |||||    ||||| || || ||||| |||||
Db    1207 AGGCCCCTTACCAGCGGGGCAGCTTAGAGAGCCAAGGGGATCTGACATAGCAGGGACAAC 1266

Qy     836 TAGTTCAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGG 895
           ||  ||||||| ||||| |||||||||||||  || | ||| || ||  ||||||||||
Db    1267 AAGCACAGTAGAAGAACAGATCCAGTGGATGTTTAGGCCACAAAATCCTGTACCAGTAGG 1326

Qy     896 CAACATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCC 955
           ||||| || || ||||||||||  | || ||||| || |||||||| ||||| |||||
Db    1327 AAACATCTATAGAAGATGGATCCAGATAGGATTGCAGAAGTGTGTCAGGATGTACAACCC 1386

Qy     956 AACAAACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAG 1015
           || ||||| |||||  ||||||| || ||||| ||||| || || |||||||||| ||
Db    1387 GACCAACATCCTAGACATAAAACAGGGACCAAAGGAGCCGTTCCAAAGCTATGTAGATAG 1446

Qy    1016 GTTCTACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCA 1075
           |||||||||| || || |||||||| |||||| ||||||| ||||||||||||||| ||
Db    1447 ATTCTACAAAAGCTTGAGGGCAGAACAAACAGATCCAGCAGTGAAGAATTGGATGACCCA 1506

Qy    1076 AACACTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGT 1135
           |||||||||  | |||||||| |||||||| || ||  ||||||| || || || ||  |
Db    1507 AACACTGCTAGTACAAAATGCCAACCCAGACTGTAAATTAGTGCTAAAAGGACTAGGGAT 1566

Qy    1136 GAATCCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAA 1195
           ||| || ||| ||||||| |||||||| || ||||| || ||||| |||| || ||||||
Db    1567 GAACCCTACCTTAGAAGAGATGCTGACCGCCTGTCAGGGGGTAGGTGGGCCAGGCCAGAA 1626

Qy    1196 GGCTAGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGC 1255
           |||||||||||||||| |||||||||||| | | ||| |  || ||||| || ||
Db    1627 AGCTAGATTAATGGCAGAGGCCCTGAAAGAGGTCATAGGACCTGCCCCTATCCCATTCGC 1686

Qy    1256 AGCAGCCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGG 1315
           ||||||||| |||          |||||| || |||| || ||||| ||||| || || ||
Db    1687 AGCAGCCCAGCAG---------AGAAAGGCATTTAAATGCTGGAACTGTGGAAAGGAAGG 1737

Qy    1316 ACACTCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAAT 1375
           ||||| ||||| |||||| |||| || ||||| ||||| |||||||| ||||| ||
Db    1738 GCACTCGGCAAGACAATGCCGAGCACCTAGAAGGCAGGGCTGCTGGAAGTGTGGTAAGCC 1797

Qy    1376 GGACCATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATG 1435
           |  ||  | ||| | || |||||||| |||||||| |||||||||| || || || ||
Db    1798 AGGACACATCATGACAAACTGCCCAGATAGACAGGCAGGTTTTTTAGGACTGGGCCCTTG 1857

Qy    1436 GGGAAAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGC 1495
           |||||||||||||||||| |||||| |||| ||||| |  |||||||||  ||||| ||
Db    1858 GGGAAAGAAGCCCCGCAACTTCCCCGTGGCCCAAGTTCCGCAGGGGCTGACACCAACAGC 1917

Qy    1496 TCCCCCAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCA 1555
           ||||||| ||| ||||| |||||||| ||  |||| || ||||||   || ||    ||
Db    1918 ACCCCCAGTGGATCCAGCAGTGGATCTACTGGAGAAATATATGCAGCAAGGGAAAAGACA 1977

Qy    1556 GAGAGA----------AAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTG---- 1601
           ||||||          | | || | |   |  || ||| | |||||  |      |
Db    1978 GAGAGAGCAGAGAGAGAGACCATACAAGGAAGTGACAGAGGACTTACTGGACCTCGAGCA 2037

Qy    1602 ----------------------------ACAGAGGATTTGCTGCACCTCAATTCTCTCTT 1633
                                        |||||||| ||||||||||||||||||||||||
Db    2038 GGGGGAGACACCATACAGGGAGCCACCAACAGAGGACTTGCTGCACCTCAATTCTCTCTT 2097

Qy    1634 TGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAGAAGTATTACTGGAT 1693
           ||||  |||||||||||||| ||  | ||||| || ||||| |||||||| ||  | ||
Db    2098 TGGAAAAGACCAGTAGTCACAGCATACATTGAGGGTCAGCCAGTAGAAGTCTTGTTAGAC 2157

Qy    1694 ACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCACATTATACCCCAAAA 1753
           |||||||||| || || || ||| ||||||||||||||||     |||||| |||||||
Db    2158 ACAGGGGCTGACGACTCAATAGTAGCAGGAATAGAGTTAGGGAACAATTATAGCCCAAAA 2217

Qy    1754 ATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAAATGTAGAAATAGAA 1813
           |||||||| |||||||| || || || ||||| || ||||| ||||||||||||||||||
Db    2218 ATAGTAGGGGGAATAGGGGGATTCATAAATACCAAGGAATATAAAAATGTAGAAATAGAA 2277

Qy    1814 GTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCCCGATTAACATTTTT 1873
           ||| ||   |||| | |    |  || || |||||||| |||||||| || |||||||||
Db    2278 GTTCTAAATAAAAAGGTACGGGCCACCATAATGACAGGCGACACCCCAATCAACATTTTT 2337

Qy    1874 GGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAAT 1912
           || |||||| | || |||||  | || |||||  |||||
Db    2338 GGCAGAAATATTCTGACAGCCTTAGGCATGTCATTAAAT 2376


RESULT 14
AAQ02830
ID   AAQ02830 standard; cDNA; 9360 BP.
XX
AC   AAQ02830;
XX
DT   29-MAY-1989  (first entry)
XX
DE   cDNA to HIV-2 RNA.
XX
KW   HIV; AIDS; Vaccine; pUC-HIV-2(GH-1).
XX
OS   HIV-2.
XX
FH   Key             Location/Qualifiers
FT   CDS             544..2109
FT                   /*tag=  a
FT                   /label=Gag reading frame
FT   CDS             1830..4934
FT                   /*tag=  b
FT                   /label=Pol gene reading frame.
FT   CDS             4867..5511
FT                   /*tag=  c
FT                   /label=Q gene reading frame
FT   CDS             5342..5677
FT                   /*tag=  d
FT                   /label=X gene reading frame
FT   CDS             5680..5994
FT                   /*tag=  e
FT                   /label=R gene reading frame
FT   CDS             6148..8703
FT                   /*tag=  f
FT                   /label=Env gene reading frame.
FT   CDS             8540..9304
FT                   /*tag=  g
FT                   /label=F gene reading frame.
XX
PN   JP01289486-A.
XX
PD   21-NOV-1989.
XX
PF   16-MAY-1988;   88JP-0119024.
XX
PR   16-MAY-1988;   88JP-0119024.
XX
PA   (TOFU ) TOA NENRYO KOGYO KK.
PA   (FUJI-) FUJI KEBIO KK.
XX
DR   WPI; 1990-005177/01.
DR   P-PSDB; AAR04024-30.
XX
PT   DNA indicating complement to RNA gene -
```

```
PT   of Human Immunodeficiency Virus type 2 used for new vaccine or
PT   diagnostic for AIDS virus.
XX
PS   Claim 2; Fig.4; 12pp; Japanese.
XX
CC   cDNA to novel HIV-2 (GH-1) has been integrated into plasmid
CC   pUC HIV-2(GH-1). Useful for diagnosis and vaccination against the virrus.
CC   Carries 7 overlapping genes in varrious reading frames, includding Gag,
CC   Pol and Env.
XX
SQ   Sequence 9360 BP; 3211 A; 1944 C; 2317 G; 1888 T; 0 other;

  Query Match           59.7%;  Score 1142.4;  DB 11;  Length 9360;
  Best Local Similarity  76.3%;  Pred. No. 8.5e-286;
  Matches 1497;  Conservative    0;  Mismatches  401;  Indels   64;  Gaps    5;

Qy      7 AGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGC--------- 57
          || |||||| |||||||||| | || |||||||||| || ||||||||||
Db    424 AGGCGGCGTGTGGAGCGGGAGTGAAACAGGCCTCCGGGTGAAGGTAAGTGCCTACACCGA 483

Qy     58 -AACACAAAAAAGAAATAGCTGTCTTTTATCCAGGAAGGGGTAATAAGATAGAGTGGGAG 116
           |||   |  |||||  |||   | |  |        |    ||  | |  |||||||
Db    484 AAACCGTAGCCAGAAAAGGCTAGTTATCCTATCTTTTAGACAGGTAGAAGATTGTGGGAG 543

Qy    117 ATGGGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGG 176
          |||||||  |||||||||||||||||  ||||||  ||||||||||||||||||||  ||||||
Db    544 ATGGGCGCGAGAAACTCCGTCTTGAGAGGGAAAAAAGCAGATGAATTAGAAAAGATTAGG 603

Qy    177 CTACGACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAA 236
          ||||  |||| ||||||||||||||||| | | || ||| | || |||||||  || |||
Db    604 TTACGGCCCAGCGGAAAGAAAAAGTACAGGCTAAAACATATTGTGTGGGCAGCGAACGAA 663

Qy    237 TTAGATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATA 296
          || || | ||| ||||| ||||| ||||||||||||   |||||||| || |||||  ||
Db    664 TTGGACAAATTCGGATTGGCAGAGAGCCTGTTGGAGTCAAAAGAAGGTTGCCAAAAGATT 723

Qy    297 CTTTCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAAT 356
          ||  | || |||| |||||||| |||||||| ||||||||||||||||||||||||| |||
Db    724 CTCACAGTTTTAGATCCATTAGTACCAACAGGGTCAGAAAATTTAAAAAGCCTTTTTAAC 783

Qy    357 ACTGTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCA 416
          || |||||||||||  |||||| | ||||||||||||||||||||| | |||||  ||||||
Db    784 ACAGTCTGCGTCATTTGGTGCTTGCACGCAGAAGAGAAAGTGAAAGATACTGAAGAAGCA 843

Qy    417 AAACAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAA 476
          |||  |  |||| |||||||| |||| ||  ||||||||||  ||||||  |||||||
Db    844 AAAAAACTAGTACAGAGACATCTAGGGGCAGAAACAGGAACTGCAGAAAAAATGCCAAGT 903

Qy    477 ACAAGTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATA 536
          |||||||||||||||||||| |||||||  |||||| |||| | ||| ||||||||  | |
Db    904 ACAAGTAGACCAACAGCACCACCTAGCGGGAGAGGAAGAAACTTCCCTGTACAACAGACA 963

Qy    537 GGTGGT---AACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAA 593
          ||  |||   |||||| ||||| |||||  | |||||  ||||  |||||||  |||||||
Db    964 GGCGGTGGCAACTATATCCACGTGCCACTGAGCCCCCGAACTCTAAATGCTTGGGTAAAA 1023

Qy    594 TTGATAGAGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAA 653
          ||  ||||||| ||||| || || |||||||||||||||||||||||| ||||| ||||||
Db   1024 TTAGTAGAGGACAAGAAGTTCGGGGCAGAAGTAGTGCCAGGATTTCAAGCACTCTCAGAA 1083

Qy    654 GGTTGCACCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCT 713
          || ||||| |||||||| || || || | |||||||||| || || ||||| |||
Db   1084 GGCTGCACGCCCTATGATATCAACCAAATGCTTAATTGTGTGGGCGATCACCAAGCAGCT 1143

Qy    714 ATGCAGATTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCA 773
          ||||| || |||||||| ||||| || || || || ||||||||||    |||||||||
Db   1144 ATGCAAATAATCAGAGAGATTATCAATGACGAAGCAGCAGATTGGGATGCACAGCACCCA 1203

Qy    774 CAACCAGCTCC---ACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGA 830
            |||||  ||   || |  ||| || ||||| || ||    || || || || ||||||
Db   1204 ATACCAGGCCCCTTACCAGCAGGGCAGCTTAGAGACCCAAGGGGGTCTGACATAGCAGGA 1263

Qy    831 ACAACTAGTTCAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCA 890
          |||||  ||  |||||||  |||||||||||| || | ||| || ||| | ||
Db   1264 ACAACAAGCACAGTAGAAGAACAGATCCAGTGGATGTATAGGCCACAAAATCCCGTGCCG 1323

Qy    891 GTAGGCAACATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTAT 950
          ||||| ||||| |||||  |||||||||||  | ||| | || || |||||||| |||||
Db   1324 GTAGGGAACATCTACAGAAGATGGATCCAGATAGGGCTACAGAAGTGTGTCAGGATGTAC 1383

Qy    951 AACCCAACAAACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTA 1010
          |||||||| |||||  |||| ||||| || || ||||| || || || ||||||||||
Db   1384 AACCCAACTAACATCTTAGACGTAAAGCAGGGACCAAAGGAACCGTTCCAGAGCTATGTG 1443

Qy   1011 GACAGGTTCTACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATG 1070
          |||||||||||  ||||| || || |||||||  |||||| | ||||||||||| ||||||
Db   1444 GACAGGTTCTATAAAAGCTTGAGGGCAGAACAAACAGATCCGGCAGTAAAGAACTGGATG 1503

Qy   1071 ACTCAAACACTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTG 1130
          || ||||| ||||| || || ||||| |||||||| |||||| |||| || || || |||
Db   1504 ACCCAAACGCTGCTAATACAGAATGCCAACCCAGACTGCAAGTTAGTACTAAAAGGACTG 1563

Qy   1131 GGTGTGAATCCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGA 1190
          ||  |||||||||||||||||| |||||||| || ||||| || ||||| || || ||
Db   1564 GGGATGAATCCCACCCTAGAAGAGATGCTGACTGCCTGTCAGGGGGTAGGTGGACCAGGC 1623

Qy   1191 CAGAAGGCTAGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCT 1250
          || || || ||| ||||||| |||||||||| |||||  |  |||||   || |||||
Db   1624 CAAAAAGCCAGACTAATGGCCGAAGCCCTGAAGGAGGCTTTGACACCACCCCCTATCCCA 1683

Qy   1251 TTTGCAGCAGCCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAA 1310
          ||||||||||||||||||         |||||||  |||||  |||||||| || || ||
Db   1684 TTTGCAGCAGCCCAACA---------AAGAAAGGTAATTAGGTGTTGGAACTGCGGAAAG 1734

Qy   1311 GAGGGACACTCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGA 1370
          |||||||||||  || || || ||| |||| || ||||||||||  ||||||||||||||
Db   1735 GAGGGACACTCGGCGAGACAGTGCCGAGCGCCTAGAAGACAGGGCTGCTGGAAATGTGGC 1794

Qy   1371 AAAATGGACCATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGT 1430
          || |  |  || || ||||| |||||||||| ||||||| |||||||||  || |||
Db   1795 AAGACAGGACACGTCATGGCAAAATGCCCAGAAAGACAGGCAGGTTTTTTAGGGATGGGT 1854

Qy   1431 CCATGGGGAAAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCA 1490
          ||||||||||||||||||||||| |||| |||| | || |||||||| |||
Db   1855 CCATGGGGAAAGAAGCCCCGCAACTTCCCCGTGGCCCAAGCTCCTCCGGGGCTGATACCA 1914

Qy   1491 ACTGCTCCCCCAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAG 1550
          || || |||||||  || ||||| ||||||||  |  | |  || ||||||  ||| |
Db   1915 ACAGCACCCCCAGCAGATCCAGCAGTGGATCTATTGGAAAGATATATGCAGCAGGGGAGA 1974

Qy   1551 CAGCAGAGAGAAAA----------------------------------------GCAGAGA 1571
           |||||||  ||  |                                        |||
Db   1975 GAGCAGAGGGAGCAGAGGGAGAGACCATACAAAGAGGTGACGGAAGACTTACTGCACCTC 2034

Qy   1572 GAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCACCTCAATTCTCTC 1631
          ||      || |    ||| ||| |||| |||||||| ||||||||||||||||||||
Db   2035 GAGCAGGGAAAAGCACCTCACAGAGAGGCGACAGAGGACTTGCTGCACCTCAATTCTCTC 2094

Qy   1632 TTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAGAAGTATTACTGG 1691
          ||||||  |||||||||||||| ||  | || || | |||||| |||||||| ||| | |
Db   2095 TTTGGAAAAGACCAGTAGTCACAGCATACATCGAGGTACAGCCGGTAGAAGTCTTATTAG 2154

Qy   1692 ATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCACATTATACCCCAA 1751
          | |||||||||||||| || || ||  |||||||| |||||||     ||||  |  ||||
Db   2155 ACACAGGGGCTGATGACTCAATAGTGGCAGGAATACAGTTAGGGGACAATTACGTTCCAA 2214

Qy   1752 AAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAAATGTAGAAATAG 1811
          ||||||| || |||||||| || || || || || |||||     |||||| |||||||||
Db   2215 AAATAGTGGGGGGAATAGGGGGATTCATAAACACCAAAGAGATTAAAAATATAGAAATAA 2274

Qy   1812 AAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCCCGATTAACATTT 1871
          ||||  ||   |||||  | | ||  || || |||||||| || ||||| || |||||| |
Db   2275 AAGTACTAAATAAAAGAGTAAGAGCCACCATAATGACAGGAGATACCCCAATCAACATCT 2334

Qy   1872 TTGGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAATT 1913
          |||| |||||| | || |||||  | || ||||||  ||||||
Db   2335 TTGGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAATT 2376
```

```
Db    2335 TTGGCAGAAATATCCTGACAGCCTTAGGCATGTCATTAAATT 2376

RESULT 15
AAC81938
ID   AAC81938 standard; DNA; 9663 BP.
XX
AC   AAC81938;
XX
DT   28-FEB-2001  (first entry)
XX
DE   Packaging vector pCM-ROD(SD36/EM) DNA fragment.
XX
KW   Encapsidation; transfer vector; nephrotropic; antiparkinsonian; anti-HIV;
KW   cytostatic; gene therapy; transgenic; retroviral packaging;
KW   gene delivery; Parkinson's disease; infectious diseases; cancer; ds.
XX
OS   Synthetic.
XX
PN   WO200040741-A2.
XX
PD   13-JUL-2000.
XX
PF   06-JAN-2000; 2000WO-US00390.
XX
PR   07-JAN-1999;   99US-0115247.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Arya SK;
XX
DR   WPI; 2000-475836/41.
XX
PT   New lentivirus transfer vector, functionally deleted for a splice donor
PT   site and comprising a packaging signal and transgene operably linked to
PT   a promoter, for improving encapsidation or transgene RNA and for gene
PT   therapy -
XX
PS   Example 8; Page 125-130; 143pp; English.
XX
CC   This invention describes a novel transfer vector derived from a
CC   lentivirus, functionally deleted for a splice donor site (SD), and
CC   comprising a packaging signal and transgene operably linked to a
CC   promoter. The products of the invention have nephrotropic,
CC   antiparkinsonian, anti-HIV, and cytostatic activity and can be used for
CC   gene therapy. Encapsidation of transgene RNA is improved using the new
CC   retroviral packaging and transfer vectors. The new transfer and packaging
CC   vectors are used as gene delivery agents and allows transfer of a
CC   transgene into the genome of non-dividing cells. They can be used to
CC   create a high-efficiency packaging cell line that provides greatly
CC   enhanced packaging of foreign DNA. Individuals suffering from a
CC   deficiency in alpha-galactosidase expression, such as Fabry disease can
CC   be treated by delivering the vectors to cells in vitro or in vivo.
CC   Parkinson's disease, infectious diseases, such as acquired
CC   immunodeficiency syndrome and cancers can be treated with the vectors.
CC   The non-infective packaging vectors can be used to detect wild-type HIV
CC   in biological samples using southern or northern blot assays. The
CC   packaging of the vector RNA is maximised, without an increase in the
CC   packaging of the viral RNA.  Deletion of sequences upstream and
CC   downstream of the 5' SD region of the HIV-2 packaging vector results in
CC   suppressed encapsidation of the packaging vector genomes without critical
CC   loss of gene expression. Functional deletion of the SD site of the
CC   transfer vector results in enhanced encapsidation of the transfer
CC   vector's genome. HIV-2 packaging vector specifically and faithfully
CC   packages its own optimally constructed transfer vector and gives better
CC   quality and titre of vector than HIV-1.
XX
SQ   Sequence 9663 BP; 3276 A; 1952 C; 2310 G; 2074 T; 51 other;

Query Match             59.2%;  Score 1132;  DB 21;  Length 9663;
Best Local Similarity   77.9%;  Pred. No. 4.3e-283;
Matches 1439;  Conservative    0;  Mismatches  355;  Indels   54;  Gaps    4;

Qy     110 GTGGGAGATGGGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAA 169
Db    1143 GTGGGAGATGGGCGCGAGAAACTCCGTCTTGAGAGGGAAAAAAGCAGATGAATTAGAAAG 1202

Qy     170 AATTAGGCTACGACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGC 229
Db    1203 AATCAGGTTACGGCCCGGCGGAAAGAAAAAGTACAGGCTAAAACATATTGTGTGGGCAGC 1262

Qy     230 AAATGAATTAGATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCA 289
Db    1263 GAATAAATTGGACAGATTCGGATTAGCAGAGAGCCTGTTGGAGTCAAAAGAGGGTTGTCA 1322

Qy     290 AAAAATACTTTCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCT 349
Db    1323 AAAAATTCTTACAGTTTTAGATCCAATGGTACCGACAGGTTCAGAAAATTTAAAAAGTCT 1382

Qy     350 TTATAATACTGTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGA 409
Db    1383 TTTTAATACTGTCTGCGTCATTTGGTGCATACACGCAGAAGAGAAAGTGAAAGATACTGA 1442

Qy     410 GGAAGCAAAACAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTAT 469
Db    1443 AGGAGCAAAACAAATAGTGCGGAGACATCTAGTGGCAGAAACAGGAACTGCAGAGAAAAT 1502

Qy     470 GCCAAAAACAAGTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACA 529
Db    1503 GCCAAGCACAAGTAGACCAACAGCACCATCTAGCGAGAAGGGAGGAAATTACCCAGTGCA 1562

Qy     530 ACAAATAGGTGGTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGT 589
Db    1563 ACATGTAGGCGGCAACTACACCCATATACCGCTGAGTCCCCGAACCCTAAATGCCTGGGT 1622

Qy     590 AAAATTGATAGAGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTC 649
Db    1623 AAAATTAGTAGAGGAAAAAAAGTTCGGGGCAGAAGTAGTGCCAGGATTTCAGGCACTCTC 1682

Qy     650 AGAAGGTTGCACCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGC 709
Db    1683 AGAAGGCTGCACGCCCTATGATATCAACCAAATGCTTAATTGTGTGGGCGACCATCAAGC 1742

Qy     710 GGCTATGCAGATTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCA 769
Db    1743 AGCCATGCAGATAATCAGGGAGATTATCAATGAGGAAGCAGCAGAATGGGATGTGCAACA 1802

Qy     770 CCCACAACCAGCTCC---ACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGC 826
Db    1803 TCCAATACCAGGCCCCTTACCAGCGGGGCAGCTTAGAGAGCCAAGGGGATCTGACATAGC 1862

Qy     827 AGGAACAACTAGTTCAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCAT 886
Db    1863 AGGGACAACAAGCACAGTAGAAGAACAGATCCAGTGGATGTTTAGGCCACAAAATCCTGT 1922

Qy     887 ACCAGTAGGCAACATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAAT 946
Db    1923 ACCAGTAGGAAACATCTATAGAAGATGGATCCAGATAGGATTGCAGAAGTGTGTCAGGAT 1982

Qy     947 GTATAACCCAACAAACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTA 1006
Db    1983 GTACAACCCGACCAACATCCTAGACATAAAACAGGGACCAAAGGAGCCGTTCCAAAGCTA 2042

Qy    1007 TGTAGACAGGTTCTACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTG 1066
Db    2043 TGTAGATAGATTCTACAAAAGCTTGAGGGCAGAACAAACAGATCCAGCAGTGAAGAATTG 2102

Qy    1067 GATGACTCAAACACTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGG 1126
Db    2103 GATGACCCAAACACTGCTAGTACAAAATGCCAACCCAGACTGTAAATTAGTGCTAAAAGG 2162

Qy    1127 GCTGGGTGTGAATCCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCC 1186
Db    2163 ACTAGGGATGAACCCTACCTTAGAAGAGATGCTGACCGCCTGTCAGGGGGTAGGTGGGCC 2222

Qy    1187 GGGACAGAAGGCTAGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAAT 1246
```

```
Db   2223 AGGCCAGAAAGCTAGATTAATGGCAGAGGCCCTGAAAGAGGTCATAGGACCTGCCCCTAT 2282

Qy   1247 CCCTTTTGCAGCAGCCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGG 1306
          ||| || |||||||||| |||         |||||| || |||| || ||||| |||||
Db   2283 CCCATTCGCAGCAGCCCAGCAG---------AGAAAGGCATTTAAATGCTGGAACTGTGG 2333

Qy   1307 GAAAGAGGGACACTCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATG 1366
          || || || ||||| ||||| |||||| |||| || ||||| ||||| |||||||| ||
Db   2334 AAAGGAAGGGCACTCGGCAAGACAATGCCGAGCACCTAGAAGGCAGGGCTGCTGGAAGTG 2393

Qy   1367 TGGAAAAATGGACCATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCT 1426
          ||| ||    |  ||  | ||| | || |||||||| |||||||| |||||||||| ||
Db   2394 TGGTAAGCCAGGACACATCATGACAAACTGCCCAGATAGACAGGCAGGTTTTTTAGGACT 2453

Qy   1427 TGGTCCATGGGGAAAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGAT 1486
          || || ||||||||||||||||||| |||||| |||| ||||| |  ||||||||||
Db   2454 GGGCCCTTGGGGAAAGAAGCCCCGCAACTTCCCCGTGGCCCAAGTTCCGCAGGGGCTGAC 2513

Qy   1487 GCCAACTGCTCCCCCAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGG 1546
          ||||| || ||||||| ||| ||||| |||||||| ||  |||| || ||||||   ||
Db   2514 ACCAACAGCACCCCCAGTGGATCCAGCAGTGGATCTACTGGAGAAATATATGCAGCAAGG 2573

Qy   1547 CAAGCAGCAGAGAGAAAA----------------------------------------GCA 1567
            ||    ||||||||  |                                        |||
Db   2574 GAAAAGACAGAGAGAGCAGAGAGAGAGACCATACAAGGAAGTGACAGAGGACTTACTGCA 2633

Qy   1568 GAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGT---GACAGAGGATTTGCTGCACCTCAA 1624
              ||     | ||||  || |||| ||||      |||||||| ||||||||||||||
Db   2634 CCTCGAGCAGGGGGAGACACCATACAGGGAGCCACCAACAGAGGACTTGCTGCACCTCAA 2693

Qy   1625 TTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAGAAGTA 1684
          |||||||||||||  |||||||||||||| ||  | ||||| || ||||| ||||||||
Db   2694 TTCTCTCTTTGGAAAAGACCAGTAGTCACAGCATACATTGAGGGTCAGCCAGTAGAAGTC 2753

Qy   1685 TTACTGGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCACATTAT 1744
          ||  | || |||||||||| || || || ||| ||||||||||||||||     |||||
Db   2754 TTGTTAGACACAGGGGCTGACGACTCAATAGTAGCAGGAATAGAGTTAGGGAACAATTAT 2813

Qy   1745 ACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAAATGTA 1804
          | ||||||||||||||| |||||||| || || || ||||| || ||||| |||||||||
Db   2814 AGCCCAAAAATAGTAGGGGGAATAGGGGGATTCATAAATACCAAGGAATATAAAAATGTA 2873

Qy   1805 GAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCCCGATT 1864
          |||||||||||| ||   |||| | |    |  || || |||||||| |||||||| ||
Db   2874 GAAATAGAAGTTCTAAATAAAAAGGTACGGGCCACCATAATGACAGGCGACACCCCAATC 2933

Qy   1865 AACATTTTTGGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAAT 1912
          |||||||||| |||||| | || |||||  | || |||||  |||||
Db   2934 AACATTTTTGGCAGAAATATTCTGACAGCCTTAGGCATGTCATTAAAT 2981


Search completed: November 16, 2002, 07:51:13
Job time : 344.38 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 16, 2002, 06:34:38 ; Search time 44.3446 Seconds
(without alignments)
13229.850 Million cell updates/sec

Title: M33262_COPY_1193_3105
Perfect score: 1913
Sequence: 1 GGTACCAGACGGCGTGAGGA.....TCTGGGGATGTCTCTAAATT 1913

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

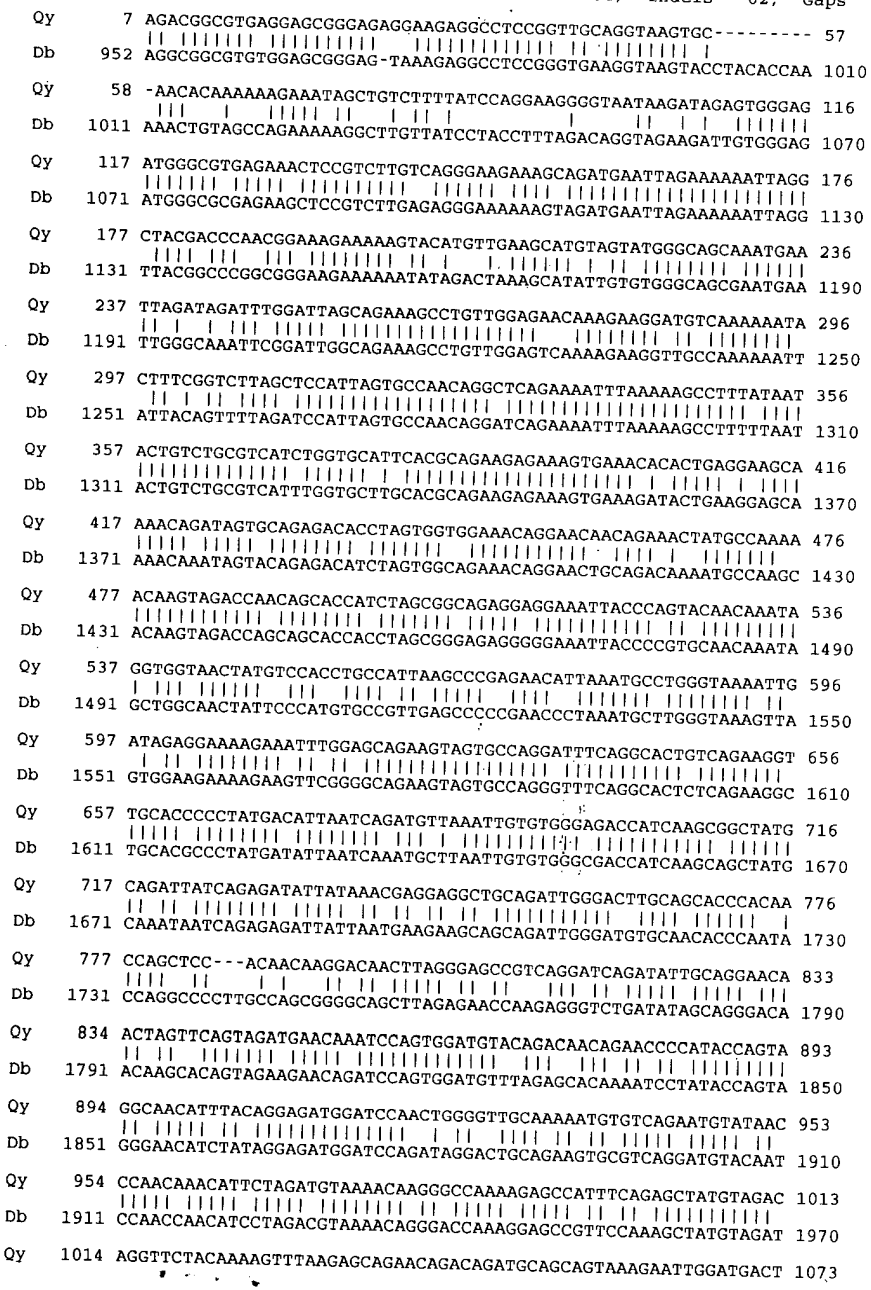
SUMMARIES

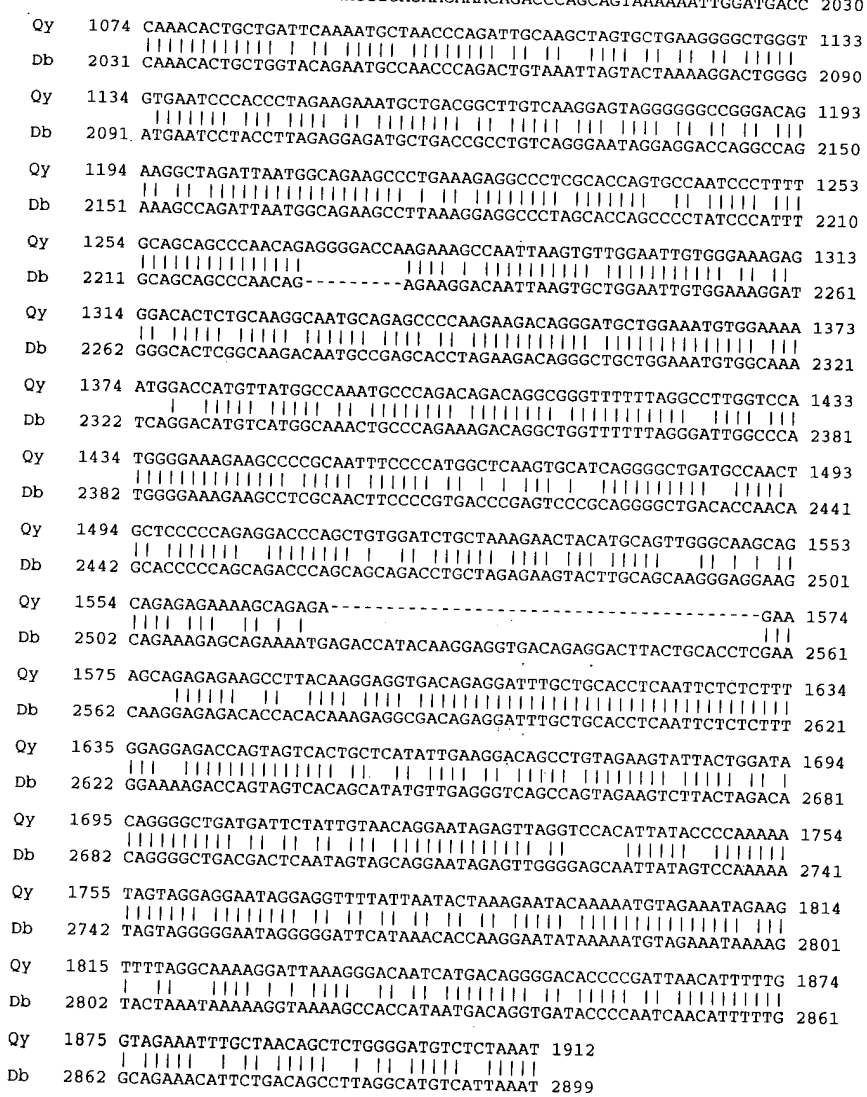
Result No.	Score	% Query Match	Length	DB	ID	Description
1	1174.4	61.4	10163	2	US-08-659-251-1	Sequence 1, Appli
2	1174.4	61.4	10163	4	US-09-256-490-1	Sequence 1, Appli
3	1174.4	61.4	10163	5	PCT-US96-11445-1	Sequence 1, Appli
4	1126.8	58.9	9633	6	5223423-1	Patent No. 5223423
5	509.6	26.6	8932	4	US-09-124-900-1	Sequence 1, Appli
6	509.6	26.6	8933	3	US-08-463-210-4	Sequence 4, Appli
7	509.6	26.6	8933	4	US-09-620-958A-3	Sequence 3, Appli
8	509.6	26.6	8933	4	US-09-620-958A-4	Sequence 4, Appli
9	509.6	26.6	8933	4	US-09-620-958A-9	Sequence 9, Appli
10	504.8	26.4	5362	3	US-08-463-210-5	Sequence 5, Appli
11	498.8	26.1	9746	1	US-08-022-835-3	Sequence 3, Appli
12	498.8	26.1	9746	1	US-08-388-809-3	Sequence 3, Appli
13	498.8	26.1	9746	2	US-08-647-714-3	Sequence 3, Appli
14	497	26.0	9737	2	US-08-944-449-7	Sequence 7, Appli
15	497	26.0	9737	4	US-09-353-362-7	Sequence 7, Appli
16	496.8	26.0	9739	1	US-08-022-835-1	Sequence 1, Appli
17	496.8	26.0	9739	1	US-08-388-809-1	Sequence 1, Appli
18	496.8	26.0	9739	2	US-08-647-714-1	Sequence 1, Appli
19	494.8	25.9	9207	3	US-08-388-353-800	Sequence 800, App
20	494.8	25.9	9207	3	US-08-488-551B-800	Sequence 800, App
21	486.6	25.4	3807	2	US-08-417-210A-78	Sequence 78, Appl
22	485.2	25.4	7399	2	US-08-418-848A-9	Sequence 9, Appli
23	484.8	25.3	12494	4	US-08-935-312-13	Sequence 13, Appl
24	484.8	25.3	12494	4	US-08-848-760B-33	Sequence 33, Appl
25	483.8	25.3	9709	2	US-08-188-583-5	Sequence 5, Appli
26	483.8	25.3	9709	3	US-08-388-353-1	Sequence 1, Appli
27	483.8	25.3	9709	3	US-08-488-551B-1	Sequence 1, Appli
28	483.8	25.3	9709	4	US-09-309-572-15	Sequence 15, Appl
29	483.8	25.3	15581	3	US-08-646-538-35	Sequence 35, Appl
30	483.8	25.3	15581	4	US-09-503-222-35	Sequence 35, Appl
31	483.2	25.3	2095	1	US-08-333-901-5	Sequence 5, Appli
32	483.2	25.3	2095	1	US-08-456-582-5	Sequence 5, Appli
33	483.2	25.3	2095	2	US-08-898-789-5	Sequence 5, Appli
34	443.6	23.2	696	4	US-09-621-625A-31	Sequence 31, Appl
35	443.4	23.2	1314	3	US-08-392-794A-1	Sequence 1, Appli
36	431.2	22.5	9793	1	US-08-470-202-56	Sequence 56, Appl
37	431.2	22.5	9793	1	US-08-471-770-56	Sequence 56, Appl
38	431.2	22.5	9793	2	US-08-468-059-56	Sequence 56, Appl
39	431.2	22.5	9793	4	US-09-109-916-56	Sequence 56, Appl
40	373.4	19.5	1733	1	US-08-470-202-58	Sequence 58, Appl
41	373.4	19.5	1733	1	US-08-471-770-58	Sequence 58, Appl
42	373.4	19.5	1733	2	US-08-468-059-58	Sequence 58, Appl
43	373.4	19.5	1733	4	US-09-109-916-58	Sequence 58, Appl
44	368.8	19.3	1733	1	US-08-470-202-57	Sequence 57, Appl
45	368.8	19.3	1733	1	US-08-471-770-57	Sequence 57, Appl

ALIGNMENTS

RESULT 1
US-08-659-251-1
; Sequence 1, Application US/08659251
; Patent No. 5883081
; GENERAL INFORMATION:
; APPLICANT: Kraus, Guenter
; APPLICANT: Wong-Staal, Flossie
; APPLICANT: Talbott, Randy
; APPLICANT: Poeschla, Eric
; TITLE OF INVENTION: Isolation of No. 5883081el HIV-2 Proviruses
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/659,251
; FILING DATE: No. 5883081 yet assigned
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/001,441
; FILING DATE: 26-JUL-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Garrett-Wackowski, Eugenia
; REGISTRATION NUMBER: 37,330
; REFERENCE/DOCKET NUMBER: 02307E-056410US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10163 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..10163
; OTHER INFORMATION: /note= "HIV-2KR provirus"
US-08-659-251-1

Query Match 61.4%; Score 1174.4; DB 2; Length 10163;
Best Local Similarity 77.4%; Pred. No. 0;
Matches 1515; Conservative 0; Mismatches 381; Indels 62; Gaps 5;





RESULT 2
US-09-256-490-1
; Sequence 1, Application US/09256490
; Patent No. 6235881
; GENERAL INFORMATION:
; APPLICANT: Kraus, Guenter
; APPLICANT: Wong-Staal, Flossie
; APPLICANT: Talbott, Randy
; APPLICANT: Poeschla, Eric

```
;    TITLE OF INVENTION:  Isolation of No. 6235881el HIV-2 Proviruses
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew LLP
;      STREET:  Two Embarcadero Center, Eighth Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/256,490
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/659,251
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Garrett-Wackowski, Eugenia
;      REGISTRATION NUMBER:  37,330
;      REFERENCE/DOCKET NUMBER:  02307E-056410US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 576-0200
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10163 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  -
;      LOCATION:  1..10163
;      OTHER INFORMATION:  /note= "HIV-2KR provirus"
US-09-256-490-1

  Query Match             61.4%;  Score 1174.4;  DB 4;  Length 10163;
  Best Local Similarity   77.4%;  Pred. No. 0;
  Matches 1515;  Conservative    0;  Mismatches  381;  Indels   62;  Gaps    5;

Qy       7 AGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGC--------- 57
           || ||||||| ||||||||||    |||||||||||| || |||||||| |
Db     952 AGGCGGCGTGTGGAGCGGGAG-TAAAGAGGCCTCCGGGTGAAGGTAAGTACCTACACCAA 1010

Qy      58 -AACACAAAAAAGAAATAGCTGTCTTTTATCCAGGAAGGGGTAATAAGATAGAGTGGGAG 116
            |||   |  ||||| ||   | || |      ||  | |  |||||||
Db    1011 AAACTGTAGCCAGAAAAAGGCTTGTTATCCTACCTTTAGACAGGTAGAAGATTGTGGGAG 1070

Qy     117 ATGGGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGG 176
           ||||||| ||||| |||||||||  |||||| |||| ||||||||||||||||||||||
Db    1071 ATGGGCGCGAGAAGCTCCGTCTTGAGAGGGAAAAAAGTAGATGAATTAGAAAAAATTAGG 1130

Qy     177 CTACGACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAA 236
           ||||  |||  ||| |||||||| || |   | |||||| | || |||||||| ||||||
Db    1131 TTACGGCCCGGCGGGAAGAAAAAATATAGACTAAAGCATATTGTGTGGGCAGCGAATGAA 1190

Qy     237 TTAGATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATA 296
           || |  | ||| ||||| ||||||||||||||||||   |||||||| || ||||||||
Db    1191 TTGGGCAAATTCGGATTGGCAGAAAGCCTGTTGGAGTCAAAAGAAGGTTGCCAAAAAATT 1250

Qy     297 CTTTCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAAT 356
           || | || |||| |||||||||||||||||| ||||||||||||||||||||||| ||||
Db    1251 ATTACAGTTTTAGATCCATTAGTGCCAACAGGATCAGAAAATTTAAAAAGCCTTTTTAAT 1310

Qy     357 ACTGTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCA 416
           |||||||||||||| |||||| | |||||||||||||||||| | ||||| | ||||
Db    1311 ACTGTCTGCGTCATTTGGTGCTTGCACGCAGAAGAGAAAGTGAAAGATACTGAAGGAGCA 1370

Qy     417 AAACAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAA 476
           ||||| ||||| |||||||| |||||||  ||||||||||  |||| |  |||||||
Db    1371 AAACAAATAGTACAGAGACATCTAGTGGCAGAAACAGGAACTGCAGACAAAATGCCAAGC 1430

Qy     477 ACAAGTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATA 536
           |||||||||||| |||||||| ||||||| ||||| |||||||||| || |||||||||
Db    1431 ACAAGTAGACCAGCAGCACCACCTAGCGGGAGAGGGGGAAATTACCCCGTGCAACAAATA 1490

Qy     537 GGTGGTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTG 596
           | ||| ||||||  |||  |||| || |||||  ||||  ||||||| |||||||| ||
Db    1491 GCTGGCAACTATTCCCATGTGCCGTTGAGCCCCCGAACCCTAAATGCTTGGGTAAAGTTA 1550

Qy     597 ATAGAGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGT 656
           | || |||||||| || || |||||||||||||||||| ||||||||||| ||||||||
Db    1551 GTGGAAGAAAAGAAGTTCGGGGCAGAAGTAGTGCCAGGGTTTCAGGCACTCTCAGAAGGC 1610

Qy     657 TGCACCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATG 716
           ||||| |||||||| |||||||| ||| | |||||||||| |||||||||| |||||||
Db    1611 TGCACGCCCTATGATATTAATCAAATGCTTAATTGTGTGGGCGACCATCAAGCAGCTATG 1670

Qy     717 CAGATTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAA 776
           || || |||||||| ||||| || || || || |||||||||||  |||| ||||||  |
Db    1671 CAAATAATCAGAGAGATTATTAATGAAGAAGCAGCAGATTGGGATGTGCAACACCCAATA 1730

Qy     777 CCAGCTCC---ACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACA 833
           ||||  ||    | |   || || |||| || ||  ||| || |||| |||| |||
Db    1731 CCAGGCCCCTTGCCAGCGGGGCAGCTTAGAGAACCAAGAGGGTCTGATATAGCAGGGACA 1790

Qy     834 ACTAGTTCAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTA 893
           || ||  ||||||| ||||| ||||||||||||  ||| ||| || || ||||||||||
Db    1791 ACAAGCACAGTAGAAGAACAGATCCAGTGGATGTTTAGAGCACAAAATCCTATACCAGTA 1850

Qy     894 GGCAACATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAAC 953
           || ||||| || |||||||||||||  | ||  |||| || || ||||| |||||| ||
Db    1851 GGGAACATCTATAGGAGATGGATCCAGATAGGACTGCAGAAGTGCGTCAGGATGTACAAT 1910

Qy     954 CCAACAAACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGAC 1013
           ||||| |||| |||| |||||||| || |||| |||| || || ||||||||||
Db    1911 CCAACCAACATCCTAGACGTAAAACAGGGACCAAAGGAGCCGTTCCAAAGCTATGTAGAT 1970

Qy     1014 AGGTTCTACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACT 1073
            || |||||||||||  |||| ||||||||  |||||  |||||||||  ||||||||||
Db    1971 AGATTCTACAAAAGCCTAAGGGCAGAACAAACAGACCCAGCAGTAAAAAATTGGATGACC 2030

Qy     1074 CAAACACTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGT 1133
            |||||||||||| | || |||||  |||||||| || ||  |||| || || || |||||
Db    2031 CAAACACTGCTGGTACAGAATGCCAACCCAGACTGTAAATTAGTACTAAAAGGACTGGGG 2090

Qy     1134 GTGAATCCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAG 1193
            ||||||| ||| |||| || |||||||| || |||| |||| |||| || || ||| |||
Db    2091 ATGAATCCTACCTTAGAGGAGATGCTGACCGCCTGTCAGGGAATAGGAGGACCAGGCCAG 2150

Qy     1194 AAGGCTAGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTT 1253
            || || |||||||||||||||||| | || |||||||| |||||||  || |||||| |||
Db    2151 AAAGCCAGATTAATGGCAGAAGCCTTAAAGGAGGCCCTAGCACCAGCCCCTATCCCATTT 2210

Qy     1254 GCAGCAGCCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAG 1313
            ||||||||||||||          |||| | |||||||||| |||||||||||| || ||
Db    2211 GCAGCAGCCCAACAG---------AGAAGGACAATTAAGTGCTGGAATTGTGGAAAGGAT 2261

Qy     1314 GGACACTCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAA 1373
            || ||||| |||| |||||| |||| || |||||||||| |||||||||||||| |||
Db    2262 GGGCACTCGGCAAGACAATGCCGAGCACCTAGAAGACAGGGCTGCTGGAAATGTGGCAAA 2321

Qy     1374 ATGGACCATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCA 1433
              |  ||||| ||||| || |||||||| |||||||| |||||||||| |||| |||
Db    2322 TCAGGACATGTCATGGCAAACTGCCCAGAAAGACAGGCTGGTTTTTTAGGGATTGGCCCA 2381

Qy     1434 TGGGGAAAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACT 1493
            |||||||||||||| ||||| |||||| || | | ||| |  |||||||||| | |||||
Db    2382 TGGGGAAAGAAGCCTCGCAACTTCCCCGTGACCCGAGTCCCGCAGGGGCTGACACCAACA 2441

Qy     1494 GCTCCCCCAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAG 1553
```

```
Db   2442 GCACCCCCAGCAGACCCAGCAGCAGACCTGCTAGAGAAGTACTTGCAGCAAGGGAGGAAG 2501

Qy   1554 CAGAGAGAAAAGCAGAGA----------------------------------------GAA 1574

Db   2502 CAGAAAGAGCAGAAAATGAGACCATACAAGGAGGTGACAGAGGACTTACTGCACCTCGAA 2561

Qy   1575 AGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCACCTCAATTCTCTCTTT 1634

Db   2562 CAAGGAGAGACACCACACAAAGAGGCGACAGAGGATTTGCTGCACCTCAATTCTCTCTTT 2621

Qy   1635 GGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTAGAAGTATTACTGGATA 1694

Db   2622 GGAAAAGACCAGTAGTCACAGCATATGTTGAGGGTCAGCCAGTAGAAGTCTTACTAGACA 2681

Qy   1695 CAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCACATTATACCCCAAAAA 1754

Db   2682 CAGGGGCTGACGACTCAATAGTAGCAGGAATAGAGTTGGGGAGCAATTATAGTCCAAAAA 2741

Qy   1755 TAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAAAATGTAGAAATAGAAG 1814

Db   2742 TAGTAGGGGGAATAGGGGGATTCATAAACACCAAGGAATATAAAAATGTAGAAATAAAAG 2801

Qy   1815 TTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACCCCGATTAACATTTTTG 1874

Db   2802 TACTAAATAAAAAGGTAAAAGCCACCATAATGACAGGTGATACCCCAATCAACATTTTTG 2861

Qy   1875 GTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAAAT 1912

Db   2862 GCAGAAACATTCTGACAGCCTTAGGCATGTCATTAAAT 2899
```

```
RESULT 3
PCT-US96-11445-1
; Sequence 1, Application PC/TUS9611445
;  GENERAL INFORMATION:
;    APPLICANT:  The Regents of the University of California
;    TITLE OF INVENTION:  Isolation of Novel HIV-2 Proviruses
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Robbins, Berliner & Carson
;      STREET:  201 N. Figueroa Street, 5th Floor
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  90012-2628
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/11445
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berliner, Robert
;      REGISTRATION NUMBER:  20,121
;      REFERENCE/DOCKET NUMBER:  5555-399C1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 977-1001
;      TELEFAX:  (213) 977-1003
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10163 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  -
;      LOCATION:  1..10163
;      OTHER INFORMATION:  /note= "HIV-2KR provirus"
```

PCT-US96-11445-1

Query Match 61.4%; Score 1174.4; DB 5; Length 10163;
Best Local Similarity 77.4%; Pred. No. 0;
Matches 1515; Conservative 0; Mismatches 381; Indels 62; Gaps 5;

```
Qy      7 AGACGGCGTGAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGC--------- 57

Db    952 AGGCGGCGTGTGGAGCGGGAG-TAAAGAGGCCTCCGGGTGAAGGTAAGTACCTACACCAA 1010

Qy     58 -AACACAAAAAAGAAATAGCTGTCTTTTATCCAGGAAGGGGTAATAAGATAGAGTGGGAG 116

Db   1011 AAACTGTAGCCAGAAAAAGGCTTGTTATCCTACCTTTAGACAGGTAGAAGATTGTGGGAG 1070

Qy    117 ATGGGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGG 176

Db   1071 ATGGGCGCGAGAAGCTCCGTCTTGAGAGGGAAAAAAGTAGATGAATTAGAAAAAATTAGG 1130

Qy    177 CTACGACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAA 236

Db   1131 TTACGGCCCGGCGGGAAGAAAAAATATAGACTAAAGCATATTGTGTGGGCAGCGAATGAA 1190

Qy    237 TTAGATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATA 296

Db   1191 TTGGGCAAATTCGGATTGGCAGAAAGCCTGTTGGAGTCAAAAGAAGGTTGCCAAAAAATT 1250

Qy    297 CTTTCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAAT 356

Db   1251 ATTACAGTTTTAGATCCATTAGTGCCAACAGGATCAGAAAATTTAAAAAGCCTTTTTAAT 1310

Qy    357 ACTGTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCA 416

Db   1311 ACTGTCTGCGTCATTTGGTGCTTGCACGCAGAAGAGAAAGTGAAAGATACTGAAGGAGCA 1370

Qy    417 AAACAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAA 476

Db   1371 AAACAAATAGTACAGAGACATCTAGTGGCAGAAACAGGAACTGCAGACAAAATGCCAAGC 1430

Qy    477 ACAAGTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATA 536

Db   1431 ACAAGTAGACCAGCAGCACCACCTAGCGGGAGAGGGGGAAATTACCCCGTGCAACAAATA 1490

Qy    537 GGTGGTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTG 596

Db   1491 GCTGGCAACTATTCCCATGTGCCGTTGAGCCCCCGAACCCTAAATGCTTGGGTAAAGTTA 1550

Qy    597 ATAGAGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGT 656

Db   1551 GTGGAAGAAAAGAAGTTCGGGGCAGAAGTAGTGCCAGGGTTTCAGGCACTCTCAGAAGGC 1610

Qy    657 TGCACCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATG 716

Db   1611 TGCACGCCCTATGATATTAATCAAATGCTTAATTGTGTGGGCGACCATCAAGCAGCTATG 1670

Qy    717 CAGATTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAA 776

Db   1671 CAAATAATCAGAGAGATTATTAATGAAGAAGCAGCAGATTGGGATGTGCAACACCCAATA 1730

Qy    777 CCAGCTCC---ACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACA 833

Db   1731 CCAGGCCCCTTGCCAGCGGGGCAGCTTAGAGAACCAAGAGGGTCTGATATAGCAGGGACA 1790

Qy    834 ACTAGTTCAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTA 893

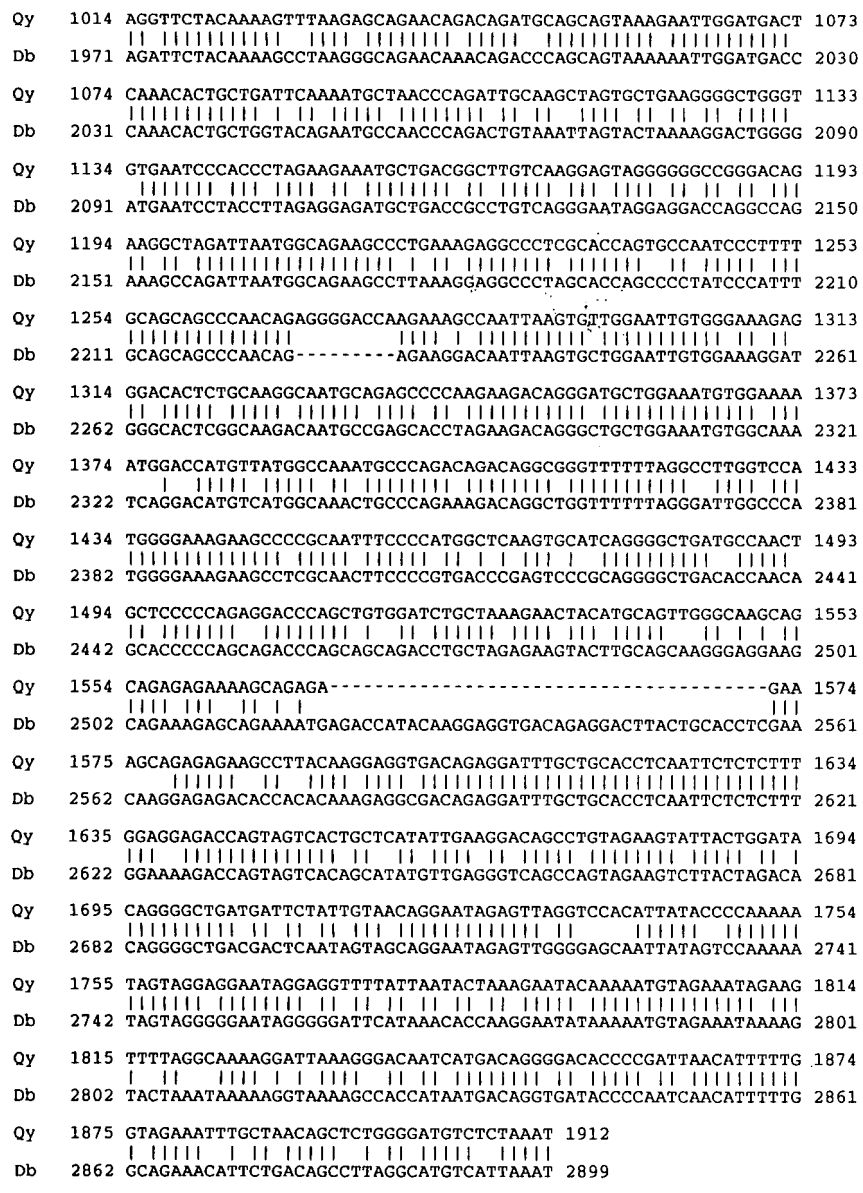
Db   1791 ACAAGCACAGTAGAAGAACAGATCCAGTGGATGTTTAGAGCACAAAATCCTATACCAGTA 1850

Qy    894 GGCAACATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAAC 953

Db   1851 GGGAACATCTATAGGAGATGGATCCAGATAGGACTGCAGAAGTGCGTCAGGATGTACAAT 1910

Qy    954 CCAACAAACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGAC 1013

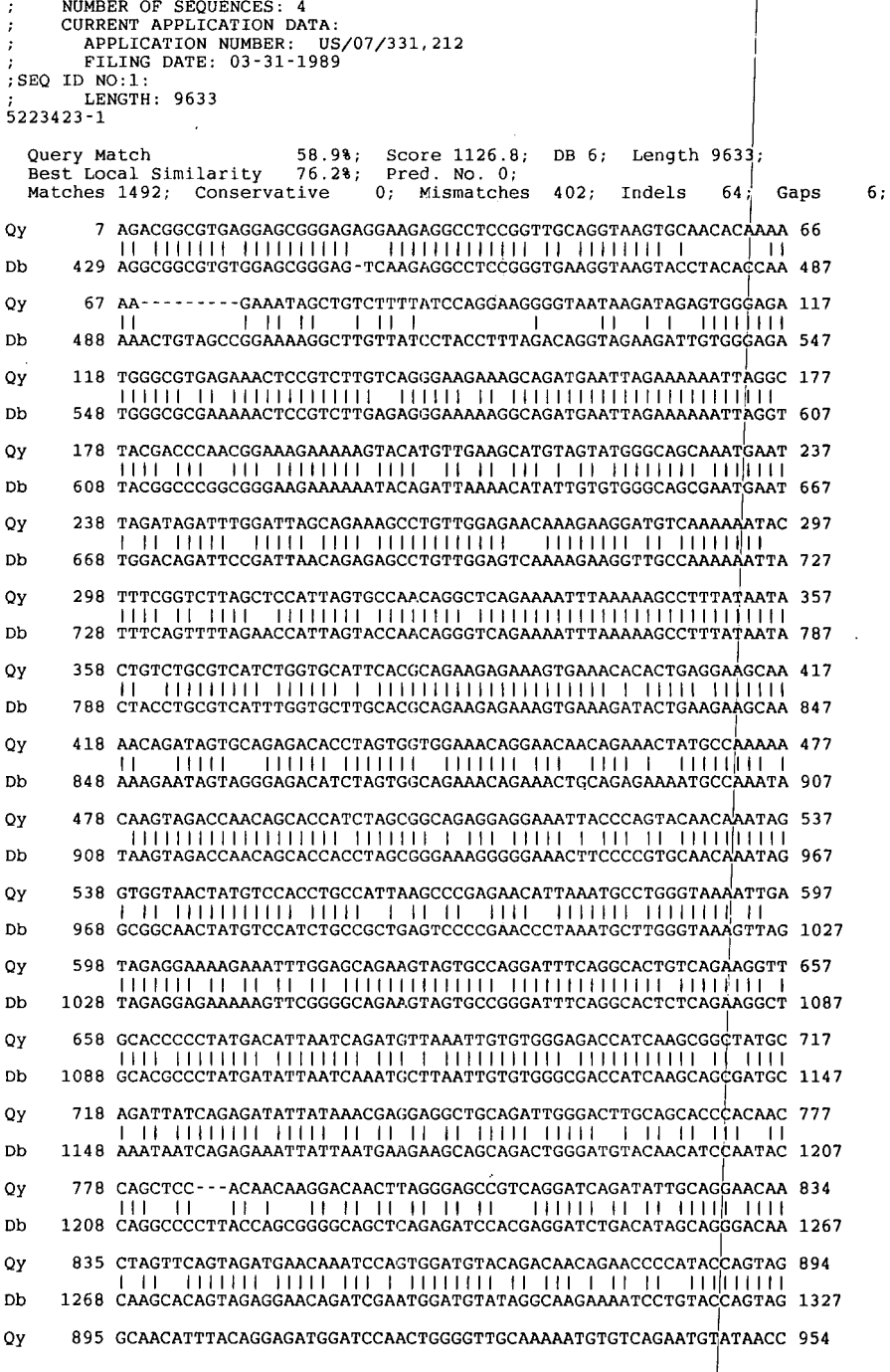
Db   1911 CCAACCAACATCCTAGACGTAAAACAGGGACCAAAGGAGCCGTTCCAAAGCTATGTAGAT 1970
```

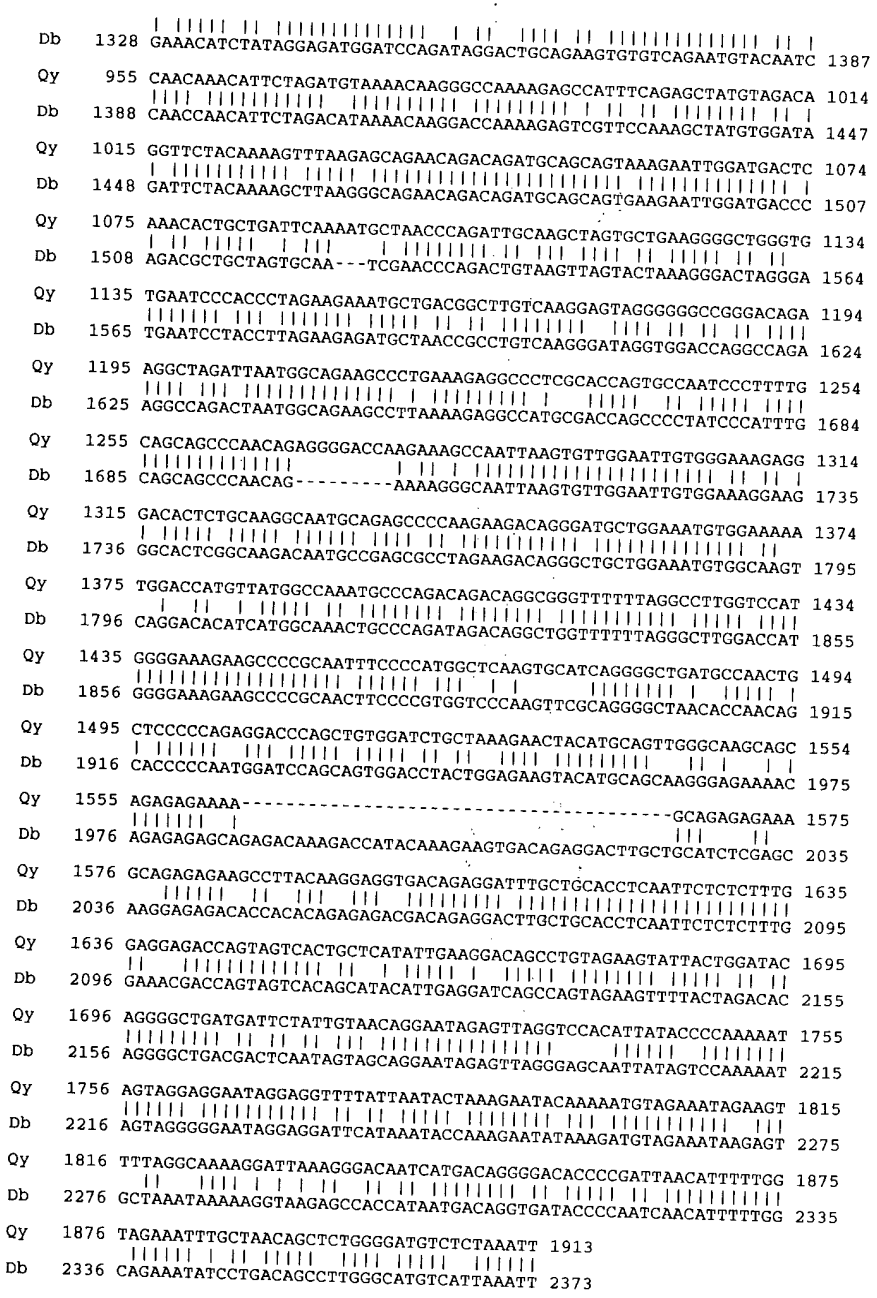


RESULT 4
5223423-1
;Patent No. 5223423
; APPLICANT: FRANCHINI, GENOVEFFA;WONG-STAAL, FLOSSIE;
;GALLO, ROBERT
; TITLE OF INVENTION: CHARACTERIZATION OF REPLICATION COMPETENT
;HUMAN IMMUNODEFICIENCY TYPE 2 PROVIRAL CLONE HIV-2 SBL/ISY

; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/331,212
; FILING DATE: 03-31-1989
;SEQ ID NO:1:
; LENGTH: 9633
5223423-1

Query Match 58.9%; Score 1126.8; DB 6; Length 9633;
Best Local Similarity 76.2%; Pred. No. 0;
Matches 1492; Conservative 0; Mismatches 402; Indels 64; Gaps 6;

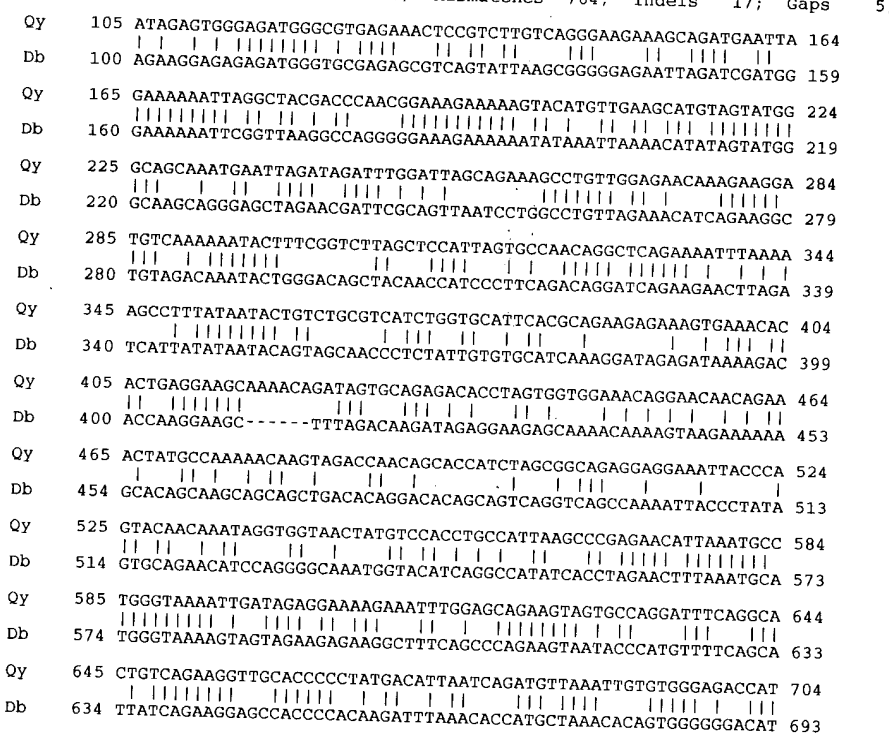




RESULT 5

US-09-124-900-1
; Sequence 1, Application US/09124900
; Patent No. 6268484
; GENERAL INFORMATION:
; APPLICANT: KATINGER, Hermann
; APPLICANT: BUCHACHER, Andreà
; APPLICANT: ERNST, Wolfgang
; APPLICANT: BALLAUN, Claudia
; APPLICANT: PURTSCHER, Martin
; APPLICANT: TRKOLA, Alexandra
; APPLICANT: PREDL, Renate
; APPLICANT: SCHMATZ, Christine
; APPLICANT: KLIMA, Annelies
; APPLICANT: STEINDL, Franz
; APPLICANT: MUSTER, Thomas
; TITLE OF INVENTION: HIV-Vaccines
; FILE REFERENCE: 1939-112P
; CURRENT APPLICATION NUMBER: US/09/124,900
; CURRENT FILING DATE: 1998-07-30
; PRIOR APPLICATION NUMBER: PCT/EP95/01481
; PRIOR FILING DATE: 1995-04-19
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 8932
; TYPE: DNA
; ORGANISM: Human imunodeficiency virus type 1
US-09-124-900-1

Query Match 26.6%; Score 509.6; DB 4; Length 8932;
Best Local Similarity 57.0%; Pred. No. 2.5e-139;
Matches 1035; Conservative 0; Mismatches 764; Indels 17; Gaps 5;



```
Qy    705 CAAGCGGCTATGCAGATTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTG 764
          ||||| || ||||| ||  | | ||| |  || || ||||| |||||||| |||||
Db    694 CAAGCAGCCATGCAAATGTTAAAAGAGACCATCAATGAGGAAGCTGCAGAATGGGATAGA 753

Qy    765 CAGCACCCACAACCAG------CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCA 818
             || |||   |  |      ||     || ||| ||  | || || ||    |||
Db    754 GTACATCCAGTGCATGCAGGGCCTATTGCACCAGGCCAGATGAGAGAACCAAGGGGAAGT 813

Qy    819 GATATTGCAGGAACAACTAGTTCAGTAGATGAACAAATCCAGTGGATGTACAGACAACAG 878
          || || |||||||| |||||| |  |  | ||||||||    ||||||   | | |  |
Db    814 GACATAGCAGGAACTACTAGTACCCTTCAGGAACAAATAGGATGGATG---ACAAATAAT 870

Qy    879 AACCCCATACCAGTAGGCAACATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGT 938
             || ||||||||  | ||||| |  ||||||||    ||||| ||  | |||
Db    871 CCACCTATCCCAGTAGGAGAAATTTATAAAAGATGGATAATCCTGGGATTAAATAAAATA 930

Qy    939 GTCAGAATGTATAACCCAACAAACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTT 998
          || |||||||||| ||| || | |||||| ||  ||| |||||| |||||||| || |||
Db    931 GTAAGAATGTATAGCCCTACCAGCATTCTGGACATAAGACAAGGACCAAAAGAACCTTTT 990

Qy    999 CAGAGCTATGTAGACAGGTTCTACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTA 1058
             |||||||||| |||||| |||| | |||||| || ||  |      |  |||
Db    991 AGAGACTATGTAGACCGGTTCTATAAAACTCTAAGAGCCGAGCAAGCTTCACAGGAGGTA 1050

Qy   1059 AAGAATTGGATGACTCAAACACTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTG 1118
          || ||||||||||  |||| || || | |||||||| |||||||||| |||     |
Db   1051 AAAAATTGGATGACAGAAACCTTGTTGGTCCAAAATGCGAACCCAGATTGTAAGACTATT 1110

Qy   1119 CTGAAGGGGCTGGGTGTGAATCCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTA 1178
           | || |   ||||        | || ||||||||||| |||| || ||||| ||||||
Db   1111 TTAAAAGCATTGGGACCAGCGGCTACACTAGAAGAAATGATGACAGCATGTCAGGGAGTA 1170

Qy   1179 GGGGGGCCGGGACAGAAGGCTAGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCA 1238
          || || || || || ||||| ||| |  |||| |||||  |||   | |           ||
Db   1171 GGAGGACCCGGCCATAAGGCAAGAGTTTTGGCTGAAGCAATGAGCCAAGTAACAAATACA 1230

Qy   1239 GTGCCAATCCCTTTTGCAGCAGCCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGG 1298
          |   | ||     |       || | |    |||   | |||||||    |||||||||
Db   1231 GCTACCATAATGATGCAGAGAGGCAATTTTAGGAACCAAAGAAAGATGGTTAAGTGTTTC 1290

Qy   1299 AATTGTGGGAAAGAGGGACACTCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGC 1358
          |||||||| ||||| || ||| | || ||  | ||||| ||||| || | | |||| ||
Db   1291 AATTGTGGCAAAGAAGGGCACACAGCCAGAAATTGCAGGGCCCCTAGGAAAAAGGGCTGT 1350

Qy   1359 TGGAAATGTGGAAAAATGGACCATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTT 1418
          ||||||||||||||    |  ||    |||    | ||  | || ||||||||   ||||
Db   1351 TGGAAATGTGGAAAGGAAGGACACCAAATGAAAGATTGTACTGAGAGACAGGCTAATTTT 1410

Qy   1419 TTAGGCCTTGGTCCATGGGGAAAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAG 1478
          |||||                | |||    |            |   |||       |
Db   1411 TTAGGGAAGATCTGGCCTTCCTACAAGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCA 1470

Qy   1479 GGGCTGATGCCAACTGCTCCCCCAGAGGACCCAGCTGTGG-ATCTGCTAAAGAACTACAT 1537
          | ||  |   |  |  |    |   ||  |  | | | |  | | ||   |  |  | |
Db   1471 GAGCCAACAGCCCCACCATTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAG 1530

Qy   1538 GCAGTTGGGCAAGCAGCAGAGAGAAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGA 1597
             |  ||   |  | ||||| || | |        || ||  || ||||     |  |
Db   1531 AGCTTCAGGTCTGGGGTAGAGACAACAACTCCCCCTCAGAAG-CAGGAGCCGATAGACAA 1589

Qy   1598 GGTGACAGAGGATTTGCTGCACCTCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCT 1657
          ||      |  |||      |||||  || ||||||||   ||||  | |||||
Db   1590 GGAACTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAACGACCCCTCGTCACAATA 1649

Qy   1658 CATATTGAAGGACAGCCTGTAGAAGTATTACTGGATACAGGGGCTGATGATTCTATTGTA 1717
           | || |  || || |     ||||   || | |||||||| || ||||||  |  |  ||
Db   1650 AAGATAGGGGGGCAACTAAAGGAAGCTCTATTAGATACAGGAGCAGATGATACAGTATTA 1709

Qy   1718 ACAGGAATAGAGTTAGGTCCACATTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTT 1777
            || |||    ||       |   |  | ||||||||  |||| ||||| |||||||||
Db   1710 GAAGAAATGAGTTTGCCAGGAAGATGGAAACCAAAAATGATAGGGGGAATTGGAGGTTTT 1769

Qy   1778 ATTAATACTAAAGAATACAAAAATGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGG 1837
          || ||    | | | ||  |  |  ||   |||||| | |  ||  | |    || |||
Db   1770 ATCAAAGTAAGACAGTATGATCAGATACTCATAGAAATCTGTGGACATAAAGCTATAGGT 1829

Qy   1838 ACAATCATGACAGGGGACACCCCGATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTG 1897
          ||| |  |   |||    || ||  | |||||  |||| |||||| || | ||    |
Db   1830 ACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAATCTGTTGACTCAGATT 1889

Qy   1898 GGGATGTCTCTAAATT 1913
          ||     || ||||||
Db   1890 GGTTGCACTTTAAATT 1905


RESULT 6
US-08-463-210-4
; Sequence 4, Application US/08463210
; Patent No. 6001977
;  GENERAL INFORMATION:
;    APPLICANT:  CHANG, Nancy T.
;    APPLICANT:  GALLO, Robert C.
;    APPLICANT:  WONG-STAAL, Flossie
;    TITLE OF INVENTION:  CLONING AND EXPRESSION OF HTLV-III DNA
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Morgan & Finnegan, L.L.P.
;      STREET:  345 Park Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10154-0053
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/463,210
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  436
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 06/693,866
;      FILING DATE:  23-JAN-1985
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 06/659,339
;      FILING DATE:  10-OCT-1984
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Serunian, Leslie A.
;      REGISTRATION NUMBER:  35,353
;      REFERENCE/DOCKET NUMBER:  2026-4193US2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 758-4800
;      TELEFAX:  (212) 751-6849
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8933 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  HTLV-III
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..8933
;      OTHER INFORMATION:  /standard_name= "Clone BH10"
;      OTHER INFORMATION:  /note= "Corresponds to nucleotide positions 222 to
;      OTHER INFORMATION:  9154 in figure 3 of EP 85307260"
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  113..1648
;      OTHER INFORMATION:  /product= "gag"
;    FEATURE:
;      NAME/KEY:  mat_peptide
```

```
;      LOCATION:  1408..4452
;      OTHER INFORMATION:  /product= "pol"
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  4367..4975
;      OTHER INFORMATION:  /product= "sor"
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  5560..8148
;      OTHER INFORMATION:  /product= "env"
US-08-463-210-4

  Query Match             26.6%;  Score 509.6;  DB 3;  Length 8933;
  Best Local Similarity   57.0%;  Pred. No. 2.5e-139;
  Matches 1035;  Conservative    0;  Mismatches  764;  Indels   17;  Gaps    5;

Qy    105 ATAGAGTGGGAGATGGGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTA 164
Db    101 AGAAGGAGAGAGATGGGTGCGAGAGCGTCAGTATTAAGCGGGGGAGAATTAGATCGATGG 160

Qy    165 GAAAAAATTAGGCTACGACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGG 224
Db    161 GAAAAAATTCGGTTAAGGCCAGGGGGAAAGAAAAAATATAAATTAAAACATATAGTATGG 220

Qy    225 GCAGCAAATGAATTAGATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGA 284
Db    221 GCAAGCAGGGAGCTAGAACGATTCGCAGTTAATCCTGGCCTGTTAGAAACATCAGAAGGC 280

Qy    285 TGTCAAAAAATACTTTCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAA 344
Db    281 TGTAGACAAATACTGGGACAGCTACAACCATCCCTTCAGACAGGATCAGAAGAACTTAGA 340

Qy    345 AGCCTTTATAATACTGTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACAC 404
Db    341 TCATTATATAATACAGTAGCAACCCTCTATTGTGTGCATCAAAGGATAGAGATAAAAGAC 400

Qy    405 ACTGAGGAAGCAAAACAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAA 464
Db    401 ACCAAGGAAGC------TTTAGACAAGATAGAGGAAGAGCAAAACAAAAGTAAGAAAAAA 454

Qy    465 ACTATGCCAAAAACAAGTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCA 524
Db    455 GCACAGCAAGCAGCAGCTGACACAGGACACAGCAGTCAGGTCAGCCAAAATTACCCTATA 514

Qy    525 GTACAACAAATAGGTGGTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCC 584
Db    515 GTGCAGAACATCCAGGGGCAAATGGTACATCAGGCCATATCACCTAGAACTTTAAATGCA 574

Qy    585 TGGGTAAAATTGATAGAGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCA 644
Db    575 TGGGTAAAAGTAGTAGAAGAGAAGGCTTTCAGCCCAGAAGTAATACCCATGTTTTCAGCA 634

Qy    645 CTGTCAGAAGGTTGCACCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCAT 704
Db    635 TTATCAGAAGGAGCCACCCCACAAGATTTAAACACCATGCTAAACACAGTGGGGGGACAT 694

Qy    705 CAAGCGGCTATGCAGATTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTG 764
Db    695 CAAGCAGCCATGCAAATGTTAAAAGAGACCATCAATGAGGAAGCTGCAGAATGGGATAGA 754

Qy    765 CAGCACCCACAACCAG------CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCA 818
Db    755 GTACATCCAGTGCATGCAGGGCCTATTGCACCAGGCCAGATGAGAGAACCAAGGGGAAGT 814

Qy    819 GATATTGCAGGAACAACTAGTTCAGTAGATGAACAAATCCAGTGGATGTACAGACAACAG 878
Db    815 GACATAGCAGGAACTACTAGTACCCTTCAGGAACAAATAGGATGGATG---ACAAATAAT 871

Qy    879 AACCCCATACCAGTAGGCAACATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGT 938
Db    872 CCACCTATCCCAGTAGGAGAAATTTATAAAAGATGGATAATCCTGGGATTAAATAAAATA 931

Qy    939 GTCAGAATGTATAACCCAACAAACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTT 998
Db    932 GTAAGAATGTATAGCCCTACCAGCATTCTGGACATAAGACAAGGACCAAAAGAACCTTTT 991

Qy    999 CAGAGCTATGTAGACAGGTTCTACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTA 1058
Db    992 AGAGACTATGTAGACCGGTTCTATAAAACTCTAAGAGCCGAGCAAGCTTCACAGGAGGTA 1051

Qy   1059 AAGAATTGGATGACTCAAACACTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTG 1118
Db   1052 AAAAATTGGATGACAGAAACCTTGTTGGTCCAAAATGCGAACCCAGATTGTAAGACTATT 1111

Qy   1119 CTGAAGGGGCTGGGTGTGAATCCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTA 1178
Db   1112 TTAAAAGCATTGGGACCAGCGGCTACACTAGAAGAAATGATGACAGCATGTCAGGGAGTA 1171

Qy   1179 GGGGGGCCGGGACAGAAGGCTAGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCA 1238
Db   1172 GGAGGACCCGGCCATAAGGCAAGAGTTTTGGCTGAAGCAATGAGCCAAGTAACAAATACA 1231

Qy   1239 GTGCCAATCCCTTTTGCAGCAGCCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGG 1298
Db   1232 GCTACCATAATGATGCAGAGAGGCAATTTTAGGAACCAAAGAAAGATGGTTAAGTGTTTC 1291

Qy   1299 AATTGTGGGAAAGAGGGACACTCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGC 1358
Db   1292 AATTGTGGCAAAGAAGGGCACACAGCCAGAAATTGCAGGGCCCCTAGGAAAAAGGGCTGT 1351

Qy   1359 TGGAAATGTGGAAAAATGGACCATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTT 1418
Db   1352 TGGAAATGTGGAAAGGAAGGACACCAAATGAAAGATTGTACTGAGAGACAGGCTAATTTT 1411

Qy   1419 TTAGGCCTTGGTCCATGGGGAAAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAG 1478
Db   1412 TTAGGGAAGATCTGGCCTTCCTACAAGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCA 1471

Qy   1479 GGGCTGATGCCAACTGCTCCCCCAGAGGACCCAGCTGTGG-ATCTGCTAAAGAACTACAT 1537
Db   1472 GAGCCAACAGCCCCACCATTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAG 1531

Qy   1538 GCAGTTGGGCAAGCAGCAGAGAGAAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGA 1597
Db   1532 AGCTTCAGGTCTGGGGTAGAGACAACAACTCCCCCTCAGAAG-CAGGAGCCGATAGACAA 1590

Qy   1598 GGTGACAGAGGATTTGCTGCACCTCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCT 1657
Db   1591 GGAACTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAACGACCCCTCGTCACAATA 1650

Qy   1658 CATATTGAAGGACAGCCTGTAGAAGTATTACTGGATACAGGGGCTGATGATTCTATTGTA 1717
Db   1651 AAGATAGGGGGGCAACTAAAGGAAGCTCTATTAGATACAGGAGCAGATGATACAGTATTA 1710

Qy   1718 ACAGGAATAGAGTTAGGTCCACATTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTT 1777
Db   1711 GAAGAAATGAGTTTGCCAGGAAGATGGAAACCAAAAATGATAGGGGGAATTGGAGGTTTT 1770

Qy   1778 ATTAATACTAAAGAATACAAAAATGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGG 1837
Db   1771 ATCAAAGTAAGACAGTATGATCAGATACTCATAGAAATCTGTGGACATAAAGCTATAGGT 1830

Qy   1838 ACAATCATGACAGGGGACACCCCGATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTG 1897
Db   1831 ACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAATCTGTTGACTCAGATT 1890

Qy   1898 GGGATGTCTCTAAATT 1913
Db   1891 GGTTGCACTTTAAATT 1906


RESULT 7
US-09-620-958A-3
; Sequence 3, Application US/09620958A
; Patent No. 6294338
; GENERAL INFORMATION:
```

```
;  APPLICANT: Nunomura, Kiyotada
;  TITLE OF INVENTION: POLYNUCLEOTIDE AMPLIFICATION METHOD
;  FILE REFERENCE: GP104-02.UT
;  CURRENT APPLICATION NUMBER: US/09/620,958A
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
;   LENGTH: 8933
;   TYPE: RNA
;   ORGANISM: Human Immunodeficiency Virus
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)...(8933)
;   OTHER INFORMATION: Sequence of transcripts produced from the BH10
; Patent No. 6294338
;   OTHER INFORMATION: plasmid.
US-09-620-958A-3

  Query Match             26.6%;  Score 509.6;  DB 4;  Length 8933;
  Best Local Similarity   45.2%;  Pred. No. 2.5e-139;
  Matches  821;  Conservative  214;  Mismatches  764;  Indels   17;  Gaps    5;

Qy    105 ATAGAGTGGGAGATGGGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTA 164
Db    101 AGAAGGAGAGAGAUGGGUGCGAGAGCGUCAGUAUUAAGCGGGGGAGAAUUAGAUCGAUGG 160

Qy    165 GAAAAAATTAGGCTACGACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGG 224
Db    161 GAAAAAAUUCGGUUAAGGCCAGGGGGAAAGAAAAAAUAUAAAUUAAAACAUAUAGUAUGG 220

Qy    225 GCAGCAAATGAATTAGATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGA 284
Db    221 GCAAGCAGGGAGCUAGAACGAUUCGCAGUUAAUCCUGGCCUGUUAGAAACAUCAGAAGGC 280

Qy    285 TGTCAAAAAATACTTTCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAA 344
Db    281 UGUAGACAAAUACUGGGACAGCUACAACCAUCCCUUCAGACAGGAUCAGAAGAACUUAGA 340

Qy    345 AGCCTTTATAATACTGTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACAC 404
Db    341 UCAUUAUAUAAUACAGUAGCAACCCUCUAUUGUGUGCAUCAAAGGAUAGAGAUAAAAGAC 400

Qy    405 ACTGAGGAAGCAAAACAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAA 464
Db    401 ACCAAGGAAGC------UUUAGACAAGAUAGAGGAAGAGCAAAACAAAAGUAAGAAAAAA 454

Qy    465 ACTATGCCAAAAACAAGTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCA 524
Db    455 GCACAGCAAGCAGCAGCUGACACAGGACACAGCAGUCAGGUCAGCCAAAAUUACCCUAUA 514

Qy    525 GTACAACAAATAGGTGGTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCC 584
Db    515 GUGCAGAACAUCCAGGGGCAAAUGGUACAUCAGGCCAUAUCACCUAGAACUUUAAAUGCA 574

Qy    585 TGGGTAAAATTGATAGAGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCA 644
Db    575 UGGGUAAAAGUAGUAGAAGAGAAGGCUUUCAGCCCAGAAGUAAUACCCAUGUUUUCAGCA 634

Qy    645 CTGTCAGAAGGTTGCACCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCAT 704
Db    635 UUAUCAGAAGGAGCCACCCCACAAGAUUUAAACACCAUGCUAAACACAGUGGGGGGACAU 694

Qy    705 CAAGCGGCTATGCAGATTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTG 764
Db    695 CAAGCAGCCAUGCAAAUGUUAAAAGAGACCAUCAAUGAGGAAGCUGCAGAAUGGGAUAGA 754

Qy    765 CAGCACCCACAACCAG------CTCCACAACAAGGAGAACTTAGGGAGCCGTCAGGATCA 818
Db    755 GUACAUCCAGUGCAUGCAGGGCCUAUUGCACCAGGCCAGAUGAGAGAACCAAGGGGAAGU 814

Qy    819 GATATTGCAGGAACAACTAGTTCAGTAGATGAACAAATCCAGTGGATGTACAGACAACAG 878

Db    815 GACAUAGCAGGAACUACUAGUACCCUUCAGGAACAAAUAGGAUGGAUG---ACAAAUAAU 871

Qy    879 AACCCCATACCAGTAGGCAACATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGT 938
Db    872 CCACCUAUCCCAGUAGGAGAAAUUUAUAAAAGAUGGAUAAUCCUGGGAUUAAAUAAAAUA 931

Qy    939 GTCAGAATGTATAACCCAACAAACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTT 998
Db    932 GUAAGAAUGUAUAGCCCUACCAGCAUUCUGGACAUAAGACAAGGACCAAAAGAACCUUUU 991

Qy    999 CAGAGCTATGTAGACAGGTTCTACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTA 1058
Db    992 AGAGACUAUGUAGACCGGUUCUAUAAAACUCUAAGAGCCGAGCAAGCUUCACAGGAGGUA 1051

Qy   1059 AAGAATTGGATGACTCAAACACTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTG 1118
Db   1052 AAAAAUUGGAUGACAGAAACCUUGUUGGUCCAAAAUGCGAACCCAGAUUGUAAGACUAUU 1111

Qy   1119 CTGAAGGGGCTGGGTGTGAATCCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTA 1178
Db   1112 UUAAAAGCAUUGGGACCAGCGGCUACACUAGAAGAAAUGAUGACAGCAUGUCAGGGAGUA 1171

Qy   1179 GGGGGGCCGGGACAGAAGGCTAGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCA 1238
Db   1172 GGAGGACCCGGCCAUAAGGCAAGAGUUUUGGCUGAAGCAAUGAGCCAAGUAACAAAUACA 1231

Qy   1239 GTGCCAATCCCTTTTGCAGCAGCCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGG 1298
Db   1232 GCUACCAUAAUGAUGCAGAGAGGCAAUUUUAGGAACCAAAGAAAGAUGGUUAAGUGUUUC 1291

Qy   1299 AATTGTGGGAAAGAGGGACACTCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGC 1358
Db   1292 AAUUGUGGCAAAGAAGGGCACACAGCCAGAAAUUGCAGGGCCCCUAGGAAAAAGGGCUGU 1351

Qy   1359 TGGAAATGTGGAAAAATGGACCATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTT 1418
Db   1352 UGGAAAUGUGGAAAGGAAGGACACCAAAUGAAAGAUUGUACUGAGAGACAGGCUAAUUUU 1411

Qy   1419 TTAGGCCTTGGTCCATGGGGAAAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAG 1478
Db   1412 UUAGGGAAGAUCUGGCCUUCCUACAAGGGAAGGCCAGGGAAUUUUCUUCAGAGCAGACCA 1471

Qy   1479 GGGCTGATGCCAACTGCTCCCCCAGAGGACCCAGCTGTGG-ATCTGCTAAAGAACTACAT 1537
Db   1472 GAGCCAACAGCCCCACCAUUUCUUCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAG 1531

Qy   1538 GCAGTTGGGCAAGCAGCAGAGAGAAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGA 1597
Db   1532 AGCUUCAGGUCUGGGGUAGAGACAACAACUCCCCCUCAGAAG-CAGGAGCCGAUAGACAA 1590

Qy   1598 GGTGACAGAGGATTTGCTGCACCTCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCT 1657
Db   1591 GGAACUGUAUCCUUUAACUUCCCUCAGAUCACUCUUUGGCAACGACCCCUCGUCACAAUA 1650

Qy   1658 CATATTGAAGGACAGCCTGTAGAAGTATTACTGGATACAGGGGCTGATGATTCTATTGTA 1717
Db   1651 AAGAUAGGGGGGCAACUAAAGGAAGCUCUAUUAGAUACAGGAGCAGAUGAUACAGUAUUA 1710

Qy   1718 ACAGGAATAGAGTTAGGTCCACATTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTT 1777
Db   1711 GAAGAAAUGAGUUUGCCAGGAAGAUGGAAACCAAAAAUGAUAGGGGGAAUUGGAGGUUUU 1770

Qy   1778 ATTAATACTAAAGAATACAAAAATGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGG 1837
Db   1771 AUCAAAGUAAGACAGUAUGAUCAGAUACUCAUAGAAAUCUGUGGACAUAAAGCUAUAGGU 1830

Qy   1838 ACAATCATGACAGGGGACACCCCGATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTG 1897
Db   1831 ACAGUAUUAGUAGGACCUACACCUGUCAACAUAAUUGGAAGAAAUCUGUUGACUCAGAUU 1890

Qy   1898 GGGATGTCTCTAAATT 1913
Db   1891 GGUUGCACUUUAAAUU 1906
```

```
RESULT 8
US-09-620-958A-4
; Sequence 4, Application US/09620958A
; Patent No. 6294338
; GENERAL INFORMATION:
;  APPLICANT: Nunomura, Kiyotada
;  TITLE OF INVENTION: POLYNUCLEOTIDE AMPLIFICATION METHOD
;  FILE REFERENCE: GP104-02.UT
;  CURRENT APPLICATION NUMBER: US/09/620,958A
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
;   LENGTH: 8933
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Sequence of the IAC-Asrc pseudo target
; Patent No. 6294338
;   NAME/KEY: mutation
;   LOCATION: (4135)...(4155)
;   OTHER INFORMATION: Mutated positions: 4135, 4140-1, 4145, 4150,
;   OTHER INFORMATION: 4152-3, 4155
US-09-620-958A-4

  Query Match             26.6%;  Score 509.6;  DB 4;  Length 8933;
  Best Local Similarity   45.2%;  Pred. No. 2.5e-139;
  Matches  821;  Conservative  214;  Mismatches  764;  Indels   17;  Gaps    5;

Qy    105 ATAGAGTGGGAGATGGGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTA 164
Db    101 AGAAGGAGAGAGAUGGGUGCGAGAGCGUCAGUAUUAAGCGGGGGAGAAUUAGAUCGAUGG 160

Qy    165 GAAAAAATTAGGCTACGACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGG 224
Db    161 GAAAAAAUUCGGUUAAGGCCAGGGGGAAAGAAAAAAUAUAAAUUAAAACAUAUAGUAUGG 220

Qy    225 GCAGCAAATGAATTAGATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGA 284
Db    221 GCAAGCAGGGAGCUAGAACGAUUCGCAGUUAAUCCUGGCCUGUUAGAAACAUCAGAAGGC 280

Qy    285 TGTCAAAAAATACTTTCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAA 344
Db    281 UGUAGACAAAUACUGGGACAGCUACAACCAUCCCUUCAGACAGGAUCAGAAGAACUUAGA 340

Qy    345 AGCCTTTATAATACTGTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACAC 404
Db    341 UCAUUAUAUAAUACAGUAGCAACCCUCUAUUGUGUGCAUCAAAGGAUAGAGAUAAAAGAC 400

Qy    405 ACTGAGGAAGCAAAACAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAA 464
Db    401 ACCAAGGAAGC------UUUAGACAAGAUAGAGGAAGAGCAAAACAAAAGUAAGAAAAAA 454

Qy    465 ACTATGCCAAAAACAAGTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCA 524
Db    455 GCACAGCAAGCAGCAGCUGACACAGGACACAGCAGUCAGGUCAGCCAAAAUUACCCUAUA 514

Qy    525 GTACAACAAATAGGTGGTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCC 584
Db    515 GUGCAGAACAUCCAGGGGCAAAUGGUACAUCAGGCCAUAUCACCUAGAACUUUAAAUGCA 574

Qy    585 TGGGTAAAATTGATAGAGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCA 644
Db    575 UGGGUAAAAGUAGUAGAAGAGAAGGCUUUCAGCCCAGAAGUAAUACCCAUGUUUUCAGCA 634

Qy    645 CTGTCAGAAGGTTGCACCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCAT 704
Db    635 UUAUCAGAAGGAGCCACCCCACAAGAUUUAAACACCAUGCUAAACACAGUGGGGGGACAU 694

Qy    705 CAAGCGGCTATGCAGATTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTG 764
```

```
Db    695 CAAGCAGCCAUGCAAAUGUUAAAAGAGACCAUCAAUGAGGAAGCUGCAGAAUGGGAUAGA 754

Qy    765 CAGCACCCACAACCAG------CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCA 818
Db    755 GUACAUCCAGUGCAUGCAGGGCCUAUUGCACCAGGCCAGAUGAGAGAACCAAGGGGAAGU 814

Qy    819 GATATTGCAGGAACAACTAGTTCAGTAGATGAACAAATCCAGTGGATGTACAGACAACAG 878
Db    815 GACAUAGCAGGAACUACUAGUACCCUUCAGGAACAAAUAGGAUGGAUG---ACAAAUAAU 871

Qy    879 AACCCCATACCAGTAGGCAACATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGT 938
Db    872 CCACCUAUCCCAGUAGGAGAAAUUUAUAAAAGAUGGAUAAUCCUGGGAUUAAAUAAAAUA 931

Qy    939 GTCAGAATGTATAACCCAACAAACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTT 998
Db    932 GUAAGAAUGUAUAGCCCUACCAGCAUUCUGGACAUAAGACAAGGACCAAAAGAACCUUUU 991

Qy    999 CAGAGCTATGTAGACAGGTTCTACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTA 1058
Db    992 AGAGACUAUGUAGACCGGUUCUAUAAAACUCUAAGAGCCGAGCAAGCUUCACAGGAGGUA 1051

Qy   1059 AAGAATTGGATGACTCAAACACTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTG 1118
Db   1052 AAAAAUUGGAUGACAGAAACCUUGUUGGUCCAAAAUGCGAACCCAGAUUGUAAGACUAUU 1111

Qy   1119 CTGAAGGGGCTGGGTGTGAATCCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTA 1178
Db   1112 UUAAAAGCAUUGGGACCAGCGGCUACACUAGAAGAAAUGAUGACAGCAUGUCAGGGAGUA 1171

Qy   1179 GGGGGGCCGGGACAGAAGGCTAGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCA 1238
Db   1172 GGAGGACCCGGCCAUAAGGCAAGAGUUUUGGCUGAAGCAAUGAGCCAAGUAACAAAUACA 1231

Qy   1239 GTGCCAATCCCTTTTGCAGCAGCCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGG 1298
Db   1232 GCUACCAUAAUGAUGCAGAGAGGCAAUUUUAGGAACCAAAGAAAGAUGGUUAAGUGUUUC 1291

Qy   1299 AATTGTGGGAAAGAGGGACACTCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGC 1358
Db   1292 AAUUGUGGCAAAGAAGGGCACACAGCCAGAAAUUGCAGGGCCCCUAGGAAAAAGGGCUGU 1351

Qy   1359 TGGAAATGTGGAAAAATGGACCATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTT 1418
Db   1352 UGGAAAUGUGGAAAGGAAGGACACCAAAUGAAAGAUUGUACUGAGAGACAGGCUAAUUUU 1411

Qy   1419 TTAGGCCTTGGTCCATGGGGAAAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAG 1478
Db   1412 UUAGGGAAGAUCUGGCCUUCCUACAAGGGAAGGCCAGGGAAUUUUCUUCAGAGCAGACCA 1471

Qy   1479 GGGCTGATGCCAACTGCTCCCCCAGAGGACCCAGCTGTGG-ATCTGCTAAAGAACTACAT 1537
Db   1472 GAGCCAACAGCCCCACCAUUUCUUCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAG 1531

Qy   1538 GCAGTTGGGCAAGCAGCAGAGAGAAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGA 1597
Db   1532 AGCUUCAGGUCUGGGGUAGAGACAACAACUCCCCCUCAGAAG-CAGGAGCCGAUAGACAA 1590

Qy   1598 GGTGACAGAGGATTTGCTGCACCTCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCT 1657
Db   1591 GGAACUGUAUCCUUUAACUUCCCUCAGAUCACUCUUUGGCAACGACCCCUCGUCACAAUA 1650

Qy   1658 CATATTGAAGGACAGCCTGTAGAAGTATTACTGGATACAGGGGCTGATGATTCTATTGTA 1717
Db   1651 AAGAUAGGGGGGCAACUAAAGGAAGCUCUAUUAGAUACAGGAGCAGAUGAUACAGUAUUA 1710

Qy   1718 ACAGGAATAGAGTTAGGTCCACATTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTT 1777
Db   1711 GAAGAAAUGAGUUUGCCAGGAAGAUGGAAACCAAAAAUGAUAGGGGGAAUUGGAGGUUUU 1770

Qy   1778 ATTAATACTAAAGAATACAAAAATGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGG 1837
Db   1771 AUCAAAGUAAGACAGUAUGAUCAGAUACUCAUAGAAAUCUGUGGACAUAAAGCUAUAGGU 1830
```

```
Qy   1838 ACAATCATGACAGGGGACACCCCGATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTG 1897
          ||| :  :   |||    || ||  : ||||:  ::|| |||||: :| : ||     :
Db   1831 ACAGUAUUAGUAGGACCUACACCUGUCAACAUAAUUGGAAGAAAUCUGUUGACUCAGAUU 1890

Qy   1898 GGGATGTCTCTAAATT 1913
          ||     |: :|||::
Db   1891 GGUUGCACUUUAAAUU 1906


RESULT 9
US-09-620-958A-9
; Sequence 9, Application US/09620958A
; Patent No. 6294338
; GENERAL INFORMATION:
;  APPLICANT: Nunomura, Kiyotada
;  TITLE OF INVENTION: POLYNUCLEOTIDE AMPLIFICATION METHOD
;  FILE REFERENCE: GP104-02.UT
;  CURRENT APPLICATION NUMBER: US/09/620,958A
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 9
;   LENGTH: 8933
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Sequence of the IAC-Bscr pseudo target
; Patent No. 6294338
;   NAME/KEY: mutation
;   LOCATION: (4140)...(4159)
;   OTHER INFORMATION: Mutated positions: 4140-42, 4145-47, 4152,
;   OTHER INFORMATION: 4156-57, 4159
US-09-620-958A-9

  Query Match             26.6%;  Score 509.6;  DB 4;  Length 8933;
  Best Local Similarity   45.2%;  Pred. No. 2.5e-139;
  Matches  821;  Conservative  214;  Mismatches  764;  Indels   17;  Gaps    5;

Qy    105 ATAGAGTGGGAGATGGGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTA 164
          | |  | | ||||:||| | ||||   :| |: ::  ..|||    ||  |||:  |:
Db    101 AGAAGGAGAGAGAUGGGUGCGAGAGCGUCAGUAUUAAGCGGGGGAGAAUUAGAUCGAUGG 160

Qy    165 GAAAAAATTAGGCTACGACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGG 224
          |||||||:: || :| | ||    ||||||||||| :| |  :: || ||: :||:|:||
Db    161 GAAAAAAUUCGGUUAAGGCCAGGGGGAAAGAAAAAAUAUAAAUUAAAACAUAUAGUAUGG 220

Qy    225 GCAGCAAATGAATTAGATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGA 284
          |||   |  ||  :|||  ||:: | | :        |||:|:: || |    ||||||
Db    221 GCAAGCAGGGAGCUAGAACGAUUCGCAGUUAAUCCUGGCCUGUUAGAAACAUCAGAAGGC 280

Qy    285 TGTCAAAAAATACTTTCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAA 344
          :|:  | |||:||:        :|   |||:   : |  ||||| :||||| |  : | |
Db    281 UGUAGACAAAUACUGGGACAGCUACAACCAUCCCUUCAGACAGGAUCAGAAGAACUUAGA 340

Qy    345 AGCCTTTATAATACTGTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACAC 404
              : :|:||:|| |:      | :|:  :|  : ||   |       |  : ||| ||
Db    341 UCAUUAUAUAAUACAGUAGCAACCCUCUAUUGUGUGCAUCAAAGGAUAGAGAUAAAAGAC 400

Qy    405 ACTGAGGAAGCAAAACAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAA 464
          ||  |||||||        :||   ||| | |·  || |    | | | | |  | | ||
Db    401 ACCAAGGAAGC------UUUAGACAAGAUAGAGGAAGAGCAAAACAAAAGUAAGAAAAAA 454

Qy    465 ACTATGCCAAAAACAAGTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCA 524
           |  || |  | ||  :    ||  |        :  | |||    |     |       |
Db    455 GCACAGCAAGCAGCAGCUGACACAGGACACAGCAGUCAGGUCAGCCAAAAUUACCCUAUA 514

Qy    525 GTACAACAAATAGGTGGTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCC 584
          |: ||  | |:    ||  |    |: || | | |  :|   || ||||| ::|||:||
Db    515 GUGCAGAACAUCCAGGGGCAAAUGGUACAUCAGGCCAUAUCACCUAGAACUUUAAAUGCA 574

Qy    585 TGGGTAAAATTGATAGAGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCA 644
          :|||:|||| :  :||| || |||   ::  |  ||||||:|. : ||    :::   |||

Db    575 UGGGUAAAAGUAGUAGAAGAGAAGGCUUUCAGCCCAGAAGUAAUACCCAUGUUUUCAGCA 634

Qy    645 CTGTCAGAAGGTTGCACCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCAT 704
          : :|||||||   ||||||  | ||  : ||  . |:| :|||    |:||| |  ||:
Db    635 UUAUCAGAAGGAGCCACCCCACAAGAUUUAAACACCAUGCUAAACACAGUGGGGGGACAU 694

Qy    705 CAAGCGGCTATGCAGATTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTG 764
          ||||| || |:||| |:  : | ||| |  |: || ||||| ||:|||| :||||
Db    695 CAAGCAGCCAUGCAAAUGUUAAAAGAGACCAUCAAUGAGGAAGCUGCAGAAUGGGAUAGA 754

Qy    765 CAGCACCCACAACCAG------CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCA 818
            || |||   |  |      |:     || ||| ||  : || || ||    |||
Db    755 GUACAUCCAGUGCAUGCAGGGCCUAUUGCACCAGGCCAGAUGAGAGAACCAAGGGGAAGU 814

Qy    819 GATATTGCAGGAACAACTAGTTCAGTAGATGAACAAATCCAGTGGATGTACAGACAACAG 878
          || |: |||||||| ||:||: |  :  | |||||||:    :|||:|   | | | | |
Db    815 GACAUAGCAGGAACUACUAGUACCCUUCAGGAACAAAUAGGAUGGAUG---ACAAAUAAU 871

Qy    879 AACCCCATACCAGTAGGCAACATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGT 938
            || |: ||||:|||  | |:::| |  |||:|||:    |:||| ::  | |||
Db    872 CCACCUAUCCCAGUAGGAGAAAUUUAUAAAAGAUGGAUAAUCCUGGGAUUAAAUAAAAUA 931

Qy    939 GTCAGAATGTATAACCCAACAAACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTT 998
          |: ||||:|:|:| ||| || | ||::|: ||  :|| |||||| |||||||| ||:::
Db    932 GUAAGAAUGUAUAGCCCUACCAGCAUUCUGGACAUAAGACAAGGACCAAAAGAACCUUUU 991

Qy    999 CAGAGCTATGTAGACAGGTTCTACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTA 1058
            |:|:|:|||| ||::|:| |||| : :|||||| || ||  |       | | |:|
Db    992 AGAGACUAUGUAGACCGGUUCUAUAAAACUCUAAGAGCCGAGCAAGCUUCACAGGAGGUA 1051

Qy   1059 AAGAATTGGATGACTCAAACACTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTG 1118
          || ||::|||:||  |||| :| :| : |||||:|| |||||||::| ||| |  :
Db   1052 AAAAAUUGGAUGACAGAAACCUUGUUGGUCCAAAAUGCGAACCCAGAUUGUAAGACUAUU 1111

Qy   1119 CTGAAGGGGCTGGGTGTGAATCCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTA 1178
           : || |   :|||        | || |:||||||||:| :||| || :|:|| ||||:|
Db   1112 UUAAAAGCAUUGGGACCAGCGGCUACACUAGAAGAAAUGAUGACAGCAUGUCAGGGAGUA 1171

Qy   1179 GGGGGGCCGGGACAGAAGGCTAGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCA 1238
          || || || || || ||||| ||| :  :||| |||||  :||  | |       | ||
Db   1172 GGAGGACCCGGCCAUAAGGCAAGAGUUUUGGCUGAAGCAAUGAGCCAAGUAACAAAUACA 1231

Qy   1239 GTGCCAATCCCTTTTGCAGCAGCCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGG 1298
          |   | |:     :       || | |   |||   | |||||    ::|||:|::
Db   1232 GCUACCAUAAUGAUGCAGAGAGGCAAUUUUAGGAACCAAAGAAAGAUGGUUAAGUGUUUC 1291

Qy   1299 AATTGTGGGAAAGAGGGACACTCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGC 1358
          ||::|:|| ||||| || ||| | || ||  | :|||| ||||| || | | |||| :|
Db   1292 AAUUGUGGCAAAGAAGGGCACACAGCCAGAAAUUGCAGGGCCCCUAGGAAAAAGGGCUGU 1351

Qy   1359 TGGAAATGTGGAAAAATGGACCATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTT 1418
          :|||||:|:|||||    | ||    |:|    | :|  | || ||||||||  |::::
Db   1352 UGGAAAUGUGGAAAGGAAGGACACCAAAUGAAAGAUUGUACUGAGAGACAGGCUAAUUUU 1411

Qy   1419 TTAGGCCTTGGTCCATGGGGAAAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAG 1478
          ::|||                 | |||     |            :   :||      |
Db   1412 UUAGGGAAGAUCUGGCCUUCCUACAAGGGAAGGCCAGGGAAUUUUCUUCAGAGCAGACCA 1471

Qy   1479 GGGCTGATGCCAACTGCTCCCCCAGAGGACCCAGCTGTGG-ATCTGCTAAAGAACTACAT 1537
          | ||  |   |  | |   |   ||  | | | |  | | ||   |  | | | |
Db   1472 GAGCCAACAGCCCCACCAUUUCUUCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAG 1531

Qy   1538 GCAGTTGGGCAAGCAGCAGAGAGAAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGA 1597
             ||   |  | |||||  || | |        || ||  || ||||    |  |
Db   1532 AGCUUCAGGUCUGGGGUAGAGACAACAACUCCCCCUCAGAAG-CAGGAGCCGAUAGACAA 1590

Qy   1598 GGTGACAGAGGATTTGCTGCACCTCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCT 1657
          ||       |   :::       ||:||  :| |:|:::||    ||||  : |:|||
Db   1591 GGAACUGUAUCCUUUAACUUCCCUCAGAUCACUCUUUGGCAACGACCCCUCGUCACAAUA 1650

Qy   1658 CATATTGAAGGACAGCCTGTAGAAGTATTACTGGATACAGGGGCTGATGATTCTATTGTA 1717
           | |: |  || || |     ||||   :| : ||:||||| || ||:||: | |:  :|
Db   1651 AAGAUAGGGGGGCAACUAAAGGAAGCUCUAUUAGAUACAGGAGCAGAUGAUACAGUAUUA 1710
```

```
Qy   1718 ACAGGAATAGAGTTAGGTCCACATTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTT 1777
Db   1711 GAAGAAAUGAGUUUGCCAGGAAGAUGGAAACCAAAAAUGAUAGGGGGAAUUGGAGGUUUU 1770

Qy   1778 ATTAATACTAAAGAATACAAAAATGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGG 1837
Db   1771 AUCAAAGUAAGACAGUAUGAUCAGAUACUCAUAGAAAUCUGUGGACAUAAAGCUAUAGGU 1830

Qy   1838 ACAATCATGACAGGGGACACCCCGATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTG 1897
Db   1831 ACAGUAUUAGUAGGACCUACACCUGUCAACAUAAUUGGAAGAAAUCUGUUGACUCAGAUU 1890

Qy   1898 GGGATGTCTCTAAATT 1913
Db   1891 GGUUGCACUUUAAAUU 1906


RESULT 10
US-08-463-210-5
; Sequence 5, Application US/08463210
; Patent No. 6001977
;  GENERAL INFORMATION:
;    APPLICANT:  CHANG, Nancy T.
;    APPLICANT:  GALLO, Robert C.
;    APPLICANT:  WONG-STAAL, Flossie
;    TITLE OF INVENTION:  CLONING AND EXPRESSION OF HTLV-III DNA
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Morgan & Finnegan, L.L.P.
;      STREET:  345 Park Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10154-0053
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/463,210
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  436
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 06/693,866
;      FILING DATE:  23-JAN-1985
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 06/659,339
;      FILING DATE:  10-OCT-1984
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Serunian, Leslie A.
;      REGISTRATION NUMBER:  35,353
;      REFERENCE/DOCKET NUMBER:  2026-4193US2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 758-4800
;      TELEFAX:  (212) 751-6849
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5362 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  HTLV-III
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..5362
;      OTHER INFORMATION:  /standard_name= "Clone BH5"
;      OTHER INFORMATION:  /note= "Corresponds to nucleotide positions 222 to
;      OTHER INFORMATION:  5585 in figure 3 of US 06/693,866 (parent)"
```

```
US-08-463-210-5

  Query Match             26.4%;  Score 504.8;  DB 3;  Length 5362;
  Best Local Similarity   56.8%;  Pred. No. 4.8e-138;
  Matches 1032;  Conservative    0;  Mismatches  767;  Indels   17;  Gaps    5;

Qy    105 ATAGAGTGGGAGATGGGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTA 164
Db     99 AGAAGGAGAGAGATGGGTGCGAGAGCGTCAGTATTAAGCGGGGGAGAATTAGATCGATGG 158

Qy    165 GAAAAAATTAGGCTACGACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGG 224
Db    159 GAAAAAATTCGGTTAAGGCCAGGGGGAAAGAAAAAATATAAATTAAAACATATAGTATGG 218

Qy    225 GCAGCAAATGAATTAGATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGA 284
Db    219 GCAAGCAGGGAGCTAGAACGATTCGCAGTTAATCCTGGCCTGTTAGAAACATCAGAAGGC 278

Qy    285 TGTCAAAAAATACTTTCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAA 344
Db    279 TGTAGACAAATACTGGGACAGCTACAACCATCCCTTCAGACAGGATCAGAAGAACTTAGA 338

Qy    345 AGCCTTTATAATACTGTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACAC 404
Db    339 TCATTATATAATACAGTAGCAACCCTCTATTGTGTGCATCAAAGGATAGAGATAAAAGAC 398

Qy    405 ACTGAGGAAGCAAAACAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAA 464
Db    399 ACCAAGGAAGC------TTTAGACAAGATAGAGGAAGAGCAAAACAAAAGTAAGAAAAAA 452

Qy    465 ACTATGCCAAAAACAAGTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCA 524
Db    453 GCACAGCAAGCAGCAGCTGACACAGGACACAGCAGTCAGGTCAGCCAAAATTACCCTATA 512

Qy    525 GTACAACAAATAGGTGGTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCC 584
Db    513 GTGCAGAACATCCAGGGGCAAATGGTACATCAGGCCATATCACCTAGAACTTTAAATGCA 572

Qy    585 TGGGTAAAATTGATAGAGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCA 644
Db    573 TGGGTAAAAGTAGTAGAAGAGAAGGCTTTCAGCCCAGAAGTGATACCCATGTTTTCAGCA 632

Qy    645 CTGTCAGAAGGTTGCACCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCAT 704
Db    633 TTATCAGAAGGAGCCACCCCACAAGATTTAAACACCATGCTAAACACAGTGGGGGGACAT 692

Qy    705 CAAGCGGCTATGCAGATTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTG 764
Db    693 CAAGCAGCCATGCAAATGTTAAAAGAGACCATCAATGAGGAAGCTGCAGAATGGGATAGA 752

Qy    765 CAGCACCCACAACCAG------CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCA 818
Db    753 GTGCATCCAGTGCATGCAGGGCCTATCGCACCAGGCCAGATGAGAGAACCAAGGGGAAGT 812

Qy    819 GATATTGCAGGAACAACTAGTTCAGTAGATGAACAAATCCAGTGGATGTACAGACAACAG 878
Db    813 GACATAGCAGGAACTACTAGTACCCTTCAGGAACAAATAGGATGGATG---ACAAATAAT 869

Qy    879 AACCCCATACCAGTAGGCAACATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGT 938
Db    870 CCACCTATCCCAGTAGGAGAAATTTATAAAAGATGGATAATCCTGGGATTAAATAAAATA 929

Qy    939 GTCAGAATGTATAACCCAACAAACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTT 998
Db    930 GTAAGGATGTATAGTCCTACCAGCATTCTGGACATAAGACAAGGACCAAAGGAACCCTTT 989

Qy    999 CAGAGCTATGTAGACAGGTTCTACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTA 1058
Db    990 AGAGACTATGTAGACCGGTTCTATAAAACTCTAAGAGCCGAGCAAGCTTCACAGGAAGTA 1049

Qy   1059 AAGAATTGGATGACTCAAACACTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTG 1118
Db   1050 AAAAATTGGATGACAGAAACCTTGTTGGTCCAAAATGCGAACCCAGATTGTAAGACTATT 1109
```

```
Qy   1119 CTGAAGGGGCTGGGTGTGAATCCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTA 1178
          | || |   ||||       | || ||||||||||||| |||| || ||||| ||||||
Db   1110 TTAAAAGCATTGGGACCAGCGGCTACACTAGAAGAAATGATGACAGCATGTCAGGGAGTA 1169

Qy   1179 GGGGGGCCGGGACAGAAGGCTAGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCA 1238
          || || || || || ||||| ||| |  |||| ||||  |||   | |          ||
Db   1170 GGAGGACCCGGCCATAAGGCAAGAGTTTTGGCTGAAGCAATGAGCCAAGTAACAAATTCA 1229

Qy   1239 GTGCCAATCCCTTTTGCAGCAGCCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGG 1298
               | ||     |   | || | |    |||   | ||||||     |||||||||
Db   1230 ACTACCATAATGATGCAAAGAGGCAATTTTAGGAACCAAAGAAAAATTGTTAAGTGTTTC 1289

Qy   1299 AATTGTGGGAAAGAGGGACACTCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGC 1358
          |||||||| ||||| || |||   |||||  | ||||  ||||| |||| | |||| ||
Db   1290 AATTGTGGCAAAGAAGGGCACATAGCAAGAAATTGCAAGGCCCCTAGAAAAAAGGGCTGT 1349

Qy   1359 TGGAAATGTGGAAAAATGGACCATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTT 1418
          |||||||||||||||    |  ||    |||    | ||  | || ||||||||   ||||
Db   1350 TGGAAATGTGGAAAGGAAGGACACCAAATGAAAGATTGTACTGAGAGACAGGCTAATTTT 1409

Qy   1419 TTAGGCCTTGGTCCATGGGGAAAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAG 1478
          |||||                 | |||    |           |   |||       |
Db   1410 TTAGGGAAGATCTGGCCTTCCTACAAGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCA 1469

Qy   1479 GGGCTGATGCCAACTGCTCCCCCAGAGGACCCAGCTGTGG-ATCTGCTAAAGAACTACAT 1537
          | ||  |   | | |    |  ||  | | | |  | | ||   |  |  | |
Db   1470 GAGCCAACAGCCCCACCATTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAG 1529

Qy   1538 GCAGTTGGGCAAGCAGCAGAGAGAAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGA 1597
            |  ||   | | ||||| || | |       ||  ||  || ||||     |  |
Db   1530 AGCTTCAGGTCTGGGGTAGAGACAACAACTCCCCCTCAGAAG-CAGGAGCCGATAGACAA 1588

Qy   1598 GGTGACAGAGGATTTGCTGCACCTCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCT 1657
          ||      |   |||      |||||  || ||||||||    ||||  | |||||
Db   1589 GGAACTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAACGACCCCTCGTCACAATA 1648

Qy   1658 CATATTGAAGGACAGCCTGTAGAAGTATTACTGGATACAGGGGCTGATGATTCTATTGTA 1717
            | || |  || || |     ||||   || | |||||||| || |||||| |  |  ||
Db   1649 AAGATAGGGGGGCAACTAAAGGAAGCTCTATTAGATACAGGAGCAGATGATACAGTATTA 1708

Qy   1718 ACAGGAATAGAGTTAGGTCCACATTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTT 1777
            || |||    ||       |   |  | |||||||  |||| ||||| |||||||||
Db   1709 GAAGAAATGAGTTTGCCAGGAAGATGGAAACCAAAAATGATAGGGGGAATTGGAGGTTTT 1768

Qy   1778 ATTAATACTAAAGAATACAAAAATGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGG 1837
          || ||    | | | ||  |  |  ||   |||||| | |  ||  | |    || |||
Db   1769 ATCAAAGTAAGACAGTATGATCAGATACTCATAGAAATCTGTGGACATAAAGCTATAGGT 1828

Qy   1838 ACAATCATGACAGGGGACACCCCGATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTG 1897
          ||| |  |   |||    || ||  | |||||  |||| |||||| || | ||     |
Db   1829 ACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAATCTGTTGACTCAGATT 1888

Qy   1898 GGGATGTCTCTAAATT 1913
          ||     || ||||||
Db   1889 GGTTGCACTTTAAATT 1904


RESULT 11
US-08-022-835-3
; Sequence 3, Application US/08022835
; Patent No. 5420030
;  GENERAL INFORMATION:
;    APPLICANT:  Reitz Jr., Marvin S.
;    APPLICANT:  Franchini, Genoveffa
;    APPLICANT:  Markham, Phillip D.
;    APPLICANT:  Gallo, Robert C.
;    APPLICANT:  Lori, Franco C.
;    APPLICANT:  Popovic, Mikulas
;    APPLICANT:  Garnter, Suzanne
;    TITLE OF INVENTION:  MOLECULAR CLONES OF HIV-1 AND USES
;    TITLE OF INVENTION:  THEREOF
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN, DARBY & CUSHMAN
;      STREET:  Eleventh Floor, 1615 L. Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/022,835
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/599,491
;      FILING DATE:  17-OCT-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Scott, Watson T.
;      REGISTRATION NUMBER:  26,581
;      REFERENCE/DOCKET NUMBER:  WTS/5683/78507/SRL
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3000
;      TELEFAX:  (202) 822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9746 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  6243..8816
US-08-022-835-3

  Query Match             26.1%;  Score 498.8;  DB 1;  Length 9746;
  Best Local Similarity   56.9%;  Pred. No. 4e-136;
  Matches 1034;  Conservative    0;  Mismatches  747;  Indels   37;  Gaps    5;

Qy     105 ATAGAGTGGGAGATGGGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTA 164
           | |  | | |||||||| | ||||   || || ||    |||    ||  |||| | ||
Db     777 AGAAGGAGAGAGATGGGTGCGAGAGCGTCGGTATTAAGCGGGGGAGAATTAGATCGATGG 836

Qy     165 GAAAAAATTAGGCTACGACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGG 224
           ||||||||| || || | ||    |||||||||| || |  || || ||||||||||||
Db     837 GAAAAAATTCGGTTAAGGCCAGGGGGAAAGAAAAAATATAAATTAAAACATGTAGTATGG 896

Qy     225 GCAGCAAATGAATTAGATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGA 284
           |||   |  ||  ||||  |||| | | |        ||||||| || |    ||||||
Db     897 GCAAGCAGGGAGCTAGAACGATTCGCAGTCAATCCTGGCCTGTTAGAAACATCAGAAGGC 956

Qy     285 TGTCAAAAAATACTTTCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAA 344
           |||  | |||||||         ||   ||||   | |  ||||| |||||| | | |||
Db     957 TGTAGACAAATACTGGGACAGCTACAACCATCCCTTCAGACAGGATCAGAAGAACTTAAA 1016

Qy     345 AGCCTTTATAATACTGTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACAC 404
               | |||||||| ||       | |||  ||  | ||   | |      |  | ||| ||
Db    1017 TCATTATATAATACAGTAGCAACCCTCTATTGTGTGCATCAAAAGATAGAGATAAAAGAC 1076

Qy     405 ACTGAGGAAGCAAAACAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAA 464
           ||  |||||||   | ||| | |  ||  | | | |          |    || | ||| |
Db    1077 ACCAAGGAAGCTTTAGAGAAAATAGAGGAAGAGCAAAACAAAAGTAAGAAAAAAGCACAG 1136

Qy     465 ACTATGCCAAAAACAAGTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCC- 523
                   |      ||   | || |      |||       |   ||||||||||
Db    1137 CAAGCAGTAGCTGACACAGGAAACAGAGGAAACAGCAGCCAAGTCAGCCAAAATTACCCC 1196

Qy     524 --AGTACAACAAATAGGTGGTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAAT 581
             ||| ||  | ||    ||  |    || || | | |  ||   || ||||| ||||||
Db    1197 ATAGTGCAGAACATCCAGGGGCAAATGGTACATCAGGCCATATCACCTAGAACTTTAAAT 1256
```

```
Qy     582 GCCTGGGTAAAATTGATAGAGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAG 641
Db    1257 GCATGGGTAAAAGTAGTAGAAGAGAAGGCTTTCAGCCCAGAAGTAATACCCATGTTTTCA 1316

Qy     642 GCACTGTCAGAAGGTTGCACCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGAC 701
Db    1317 GCATTATCAGAAGGAGCCACCCCACAAGATTTAAACACCATGCTAAACACAGTGGGGGGA 1376

Qy     702 CATCAAGCGGCTATGCAGATTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGAC 761
Db    1377 CATCAAGCAGCCATGCAAATGTTAAAAGAGACCATCAATGAGGAAGCTGCAGAATGGGAT 1436

Qy     762 TTGCAGCACCCACAACCAG------CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGA 815
Db    1437 AGATTGCATCCAGTGCATGCAGGGCCTATTGCACCAGGCCAGATGAGAGAACCAAGGGGA 1496

Qy     816 TCAGATATTGCAGGAACAACTAGTTCAGTAGATGAACAAATCCAGTGGATGTACAGACAA 875
Db    1497 AGTGACATAGCAGGAACTACTAGTACCCTTCAGGAACAAATAGGATGGATG---ACAAAT 1553

Qy     876 CAGAACCCCATACCAGTAGGCAACATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAA 935
Db    1554 AATCCACCTATCCCAGTAGGAGAAATCTATAAAAGATGGATAATCCTGGGATTAAATAAA 1613

Qy     936 TGTGTCAGAATGTATAACCCAACAAACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCA 995
Db    1614 ATAGTAAGGATGTATAGCCCTTCCAGCATTCTGGACATAAGACAAGGACCAAAGGAACCC 1673

Qy     996 TTTCAGAGCTATGTAGACAGGTTCTACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCA 1055
Db    1674 TTTAGAGACTATGTAGACCGGTTCTATAAAACTCTAAGAGCCGAGCAAGCTTCACAGGAG 1733

Qy    1056 GTAAAGAATTGGATGACTCAAACACTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTA 1115
Db    1734 GTAAAAAATTGGATGACAGAAACCTTGTTGGTCCAAAATGCGAACCCAGATTGTAAGACT 1793

Qy    1116 GTGCTGAAGGGGCTGGGTGTGAATCCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGA 1175
Db    1794 ATTTTAAAAGCATTGGGACCAGCAGCTACACTAGAAGAAATGATGACAGCATGTCAGGGA 1853

Qy    1176 GTAGGGGGGCCGGGACAGAAGGCTAGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCA 1235
Db    1854 GTGGGAGGACCTGGTCATAAAGCAAGAGTTTTGGCGGAAGCGATGAGCCAAGTAACAAAT 1913

Qy    1236 CCAGTGCCAATCCCTTTTGCAGCAGCCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGT 1295
Db    1914 TCAGCTACCATAATGATGCAGAGAGGCAATTTTAGGAATCAAAGAAAGATTATCAAGTGC 1973

Qy    1296 TGGAATTGTGGGAAAGAGGGACACTCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGA 1355
Db    1974 TTCAATTGTGGCAAAGAAGGGCACATAGCCAAAAATTGCAGGGCCCCTAGGAAAAGGGGC 2033

Qy    1356 TGCTGGAAATGTGGAAAAATGGACCATGTTATGGCCAAATGCCCAGACAGACAGGCGGGT 1415
Db    2034 TGTTGGAAATGTGGAAAGGAAGGACACCAAATGAAAGATTGTACTGAGAGACAGGCTAAT 2093

Qy    1416 TTTTTAGGCCTTGGTCCATGGGGAAAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCAT 1475
Db    2094 TTTTTAGGGAAGATCTGGCCTTCCTGCAAGGGAAGGCAGGGAATTTTCCTCAGAGCAGAA 2153

Qy    1476 CAGGGGCTGATGCCAACTGCTCCCCCAGAGGACCCAGCTGTGGATCTGCTAAAGAACTAC 1535
Db    2154 CAG-------AGCCAACAGCCCCACCAGAAGAGAGCTTCAGGTTTGGGGAAGAGACAACA 2206

Qy    1536 ATGCAGTTGGGCAAGCAGCAGAGAGAAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAG 1595
Db    2207 ACTCCCTATCAGAAGCAGGAGAAGAAGCAGGAGACGATAGACA---------------- 2249

Qy    1596 GAGGTGACAGAGGATTTGCTGCACCTCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTG 1655
Db    2250 -AGGACCTGTATCCTTTAGCTTCCCTCAAATCACTCTTTGGCAACGACCCATTGTCACAA 2308

Qy    1656 CTCATATTGAAGGACAGCCTGTAGAAGTATTACTGGATACAGGGGCTGATGATTCTATTG 1715
Db    2309 TAAAGATAGGGGGGCAACTAAAGGAAGCTCTATTAGATACAGGAGCAGATGATACAGTAT 2368

Qy    1716 TAACAGGAATAGAGTTAGGTCCACATTATACCCCAAAAATAGTAGGAGGAATAGGAGGTT 1775
Db    2369 TAGAAGAAATGAATTTGCCAGGAAGATGGAAACCAAAAATGATAGGGGGAATTGGAGGTT 2428

Qy    1776 TTATTAATACTAAAGAATACAAAAATGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAG 1835
Db    2429 TTATCAAAGTAAGACAGTATGATCAGATAACCATAGAAATCTGTGGACATAAAGCTATAG 2488

Qy    1836 GGACAATCATGACAGGGGACACCCCGATTAACATTTTTGGTAGAAATTTGCTAACAGCTC 1895
Db    2489 GTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAATCTGTTGACTCAGC 2548

Qy    1896 TGGGGATGTCTCTAAATT 1913
Db    2549 TTGGGTGCACTTTAAATT 2566
```

RESULT 12
US-08-388-809-3
; Sequence 3, Application US/08388809
; Patent No. 5576000
; GENERAL INFORMATION:
; APPLICANT: REITZ, JR., MARVIN S.; FRANCHINI,
; APPLICANT: GENOVEFFA; MARKHAM, PHILIP D. GALLO, ROBERT
; APPLICANT: C.; LORI, FRANCO C.; POPOVIC, MIKULAS; AND
; APPLICANT: GARTNER, SUZANNE
; TITLE OF INVENTION: MOLECULAR CLONES OF HIV-1
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK, 3.5"
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/388,809
; FILING DATE: 15-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/022,835
; FILING DATE: 25-FEB-1993
; APPLICATION NUMBER: US 07/599,491
; FILING DATE: 17-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: LESLIE A. SERUNIAN
; REGISTRATION NUMBER: 35,353
; REFERENCE/DOCKET NUMBER: 2026-4092US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9746 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 6243..8816
US-08-388-809-3

Query Match 26.1%; Score 498.8; DB 1; Length 9746;

```
  Best Local Similarity   56.9%;  Pred. No. 4e-136;
  Matches 1034;  Conservative    0;  Mismatches  747;  Indels   37;  Gaps    5;

Qy     105 ATAGAGTGGGAGATGGGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTA 164
Db     777 AGAAGGAGAGAGATGGGTGCGAGAGCGTCGGTATTAAGCGGGGGAGAATTAGATCGATGG 836

Qy     165 GAAAAAATTAGGCTACGACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGG 224
Db     837 GAAAAAATTCGGTTAAGGCCAGGGGGAAAGAAAAAATATAAATTAAAACATGTAGTATGG 896

Qy     225 GCAGCAAATGAATTAGATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGA 284
Db     897 GCAAGCAGGGAGCTAGAACGATTCGCAGTCAATCCTGGCCTGTTAGAAACATCAGAAGGC 956

Qy     285 TGTCAAAAAATACTTTCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAA 344
Db     957 TGTAGACAAATACTGGGACAGCTACAACCATCCCTTCAGACAGGATCAGAAGAACTTAAA 1016

Qy     345 AGCCTTTATAATACTGTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACAC 404
Db    1017 TCATTATATAATACAGTAGCAACCCTCTATTGTGTGCATCAAAAGATAGAGATAAAAGAC 1076

Qy     405 ACTGAGGAAGCAAAACAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAA 464
Db    1077 ACCAAGGAAGCTTTAGAGAAAATAGAGGAAGAGCAAAACAAAAGTAAGAAAAAAGCACAG 1136

Qy     465 ACTATGCCAAAAACAAGTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCC- 523
Db    1137 CAAGCAGTAGCTGACACAGGAAACAGAGGAAACAGCAGCCAAGTCAGCCAAAATTACCCC 1196

Qy     524 --AGTACAACAAATAGGTGGTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAAT 581
Db    1197 ATAGTGCAGAACATCCAGGGGCAAATGGTACATCAGGCCATATCACCTAGAACTTTAAAT 1256

Qy     582 GCCTGGGTAAAATTGATAGAGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAG 641
Db    1257 GCATGGGTAAAAGTAGTAGAAGAGAAGGCTTTCAGCCCAGAAGTAATACCCATGTTTTCA 1316

Qy     642 GCACTGTCAGAAGGTTGCACCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGAC 701
Db    1317 GCATTATCAGAAGGAGCCACCCCACAAGATTTAAACACCATGCTAAACACAGTGGGGGGA 1376

Qy     702 CATCAAGCGGCTATGCAGATTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGAC 761
Db    1377 CATCAAGCAGCCATGCAAATGTTAAAAGAGACCATCAATGAGGAAGCTGCAGAATGGGAT 1436

Qy     762 TTGCAGCACCCACAACCAG------CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGA 815
Db    1437 AGATTGCATCCAGTGCATGCAGGGCCTATTGCACCAGGCCAGATGAGAGAACCAAGGGGA 1496

Qy     816 TCAGATATTGCAGGAACAACTAGTTCAGTAGATGAACAAATCCAGTGGATGTACAGACAA 875
Db    1497 AGTGACATAGCAGGAACTACTAGTACCCTTCAGGAACAAATAGGATGGATG---ACAAAT 1553

Qy     876 CAGAACCCCATACCAGTAGGCAACATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAA 935
Db    1554 AATCCACCTATCCCAGTAGGAGAAATCTATAAAAGATGGATAATCCTGGGATTAAATAAA 1613

Qy     936 TGTGTCAGAATGTATAACCCAACAAACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCA 995
Db    1614 ATAGTAAGGATGTATAGCCCTTCCAGCATTCTGGACATAAGACAAGGACCAAAGGAACCC 1673

Qy     996 TTTCAGAGCTATGTAGACAGGTTCTACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCA 1055
Db    1674 TTTAGAGACTATGTAGACCGGTTCTATAAAACTCTAAGAGCCGAGCAAGCTTCACAGGAG 1733

Qy    1056 GTAAAGAATTGGATGACTCAAACACTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTA 1115
Db    1734 GTAAAAAATTGGATGACAGAAACCTTGTTGGTCCAAAATGCGAACCCAGATTGTAAGACT 1793

Qy    1116 GTGCTGAAGGGGCTGGGTGTGAATCCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGA 1175
Db    1794 ATTTTAAAAGCATTGGGACCAGCAGCTACACTAGAAGAAATGATGACAGCATGTCAGGGA 1853

Qy    1176 GTAGGGGGGCCGGGACAGAAGGCTAGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCA 1235
Db    1854 GTGGGAGGACCTGGTCATAAAGCAAGAGTTTTGGCGGAAGCGATGAGCCAAGTAACAAAT 1913

Qy    1236 CCAGTGCCAATCCCTTTTGCAGCAGCCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGT 1295
Db    1914 TCAGCTACCATAATGATGCAGAGAGGCAATTTTAGGAATCAAAGAAAGATTATCAAGTGC 1973

Qy    1296 TGGAATTGTGGGAAAGAGGGACACTCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGA 1355
Db    1974 TTCAATTGTGGCAAAGAAGGGCACATAGCCAAAAATTGCAGGGCCCCTAGGAAAAGGGGC 2033

Qy    1356 TGCTGGAAATGTGGAAAAATGGACCATGTTATGGCCAAATGCCCAGACAGACAGGCGGGT 1415
Db    2034 TGTTGGAAATGTGGAAAGGAAGGACACCAAATGAAAGATTGTACTGAGAGACAGGCTAAT 2093

Qy    1416 TTTTTAGGCCTTGGTCCATGGGGAAAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCAT 1475
Db    2094 TTTTTAGGGAAGATCTGGCCTTCCTGCAAGGGAAGGCAGGGAATTTTCCTCAGAGCAGAA 2153

Qy    1476 CAGGGGCTGATGCCAACTGCTCCCCCAGAGGACCCAGCTGTGGATCTGCTAAAGAACTAC 1535
Db    2154 CAG-------AGCCAACAGCCCCACCAGAAGAGAGCTTCAGGTTTGGGGAAGAGACAACA 2206

Qy    1536 ATGCAGTTGGGCAAGCAGCAGAGAGAAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAG 1595
Db    2207 ACTCCCTATCAGAAGCAGGAGAAGAAGCAGGAGACGATAGACA---------------- 2249

Qy    1596 GAGGTGACAGAGGATTTGCTGCACCTCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTG 1655
Db    2250 -AGGACCTGTATCCTTTAGCTTCCCTCAAATCACTCTTTGGCAACGACCCATTGTCACAA 2308

Qy    1656 CTCATATTGAAGGACAGCCTGTAGAAGTATTACTGGATACAGGGGCTGATGATTCTATTG 1715
Db    2309 TAAAGATAGGGGGGCAACTAAAGGAAGCTCTATTAGATACAGGAGCAGATGATACAGTAT 2368

Qy    1716 TAACAGGAATAGAGTTAGGTCCACATTATACCCCAAAAATAGTAGGAGGAATAGGAGGTT 1775
Db    2369 TAGAAGAAATGAATTTGCCAGGAAGATGGAAACCAAAAATGATAGGGGGAATTGGAGGTT 2428

Qy    1776 TTATTAATACTAAAGAATACAAAAATGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAG 1835
Db    2429 TTATCAAAGTAAGACAGTATGATCAGATAACCATAGAAATCTGTGGACATAAAGCTATAG 2488

Qy    1836 GGACAATCATGACAGGGGACACCCCGATTAACATTTTTGGTAGAAATTTGCTAACAGCTC 1895
Db    2489 GTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAATCTGTTGACTCAGC 2548

Qy    1896 TGGGGATGTCTCTAAATT 1913
Db    2549 TTGGGTGCACTTTAAATT 2566


RESULT 13
US-08-647-714-3
; Sequence 3, Application US/08647714
; Patent No. 5869313
;  GENERAL INFORMATION:
;    APPLICANT:  REITZ, JR., MARVIN S.; FRANCHINI,
;    APPLICANT:  GENOVEFFA; MARKHAM, PHILIP D. GALLO, ROBERT
;    APPLICANT:  C.; LORI, FRANCO C.; POPOVIC, MIKULAS; AND
;    APPLICANT:  GARTNER, SUZANNE
;    TITLE OF INVENTION:  MOLECULAR CLONES OF HIV-1
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN
;      STREET:  345 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
```

```
;       ZIP:  10154
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  FLOPPY DISK, 3.5"
;       COMPUTER:  IBM PC COMPATIBLE
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  WORDPERFECT 5.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/647,714
;       FILING DATE:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/388,809
;       FILING DATE:  15-FEB-1995
;       APPLICATION NUMBER:  US 08/022,835
;       FILING DATE:  25-FEB-1993
;       APPLICATION NUMBER:  US 07/599,491
;       FILING DATE:  17-OCT-1990
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  LESLIE A. SERUNIAN
;       REGISTRATION NUMBER:  35,353
;       REFERENCE/DOCKET NUMBER:  2026-4092US2
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 758-4800
;       TELEFAX:  (212) 751-6849
;       TELEX:  421792
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  9746 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  6243..8816
US-08-647-714-3

  Query Match             26.1%;  Score 498.8;  DB 2;  Length 9746;
  Best Local Similarity   56.9%;  Pred. No. 4e-136;
  Matches 1034;  Conservative    0;  Mismatches  747;  Indels   37;  Gaps    5;

Qy     105 ATAGAGTGGGAGATGGGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTA 164
           | |  | | |||||||| | ||||   || || ||    |||    ||  ||||  ||
Db     777 AGAAGGAGAGAGATGGGTGCGAGAGCGTCGGTATTAAGCGGGGGAGAATTAGATCGATGG 836

Qy     165 GAAAAAATTAGGCTACGACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGG 224
           |||||||||| || || | ||    ||||||||||| || |  || || ||||||||||||
Db     837 GAAAAAATTCGGTTAAGGCCAGGGGGAAAGAAAAAATATAAATTAAAACATGTAGTATGG 896

Qy     225 GCAGCAAATGAATTAGATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGA 284
           |||   |  ||  ||||  |||| | | |       ||||||| || |    ||||||
Db     897 GCAAGCAGGGAGCTAGAACGATTCGCAGTCAATCCTGGCCTGTTAGAAACATCAGAAGGC 956

Qy     285 TGTCAAAAAATACTTTCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAA 344
           |||  | |||||||         ||   ||||   | |  ||||| |||||| |  | |||
Db     957 TGTAGACAAATACTGGGACAGCTACAACCATCCCTTCAGACAGGATCAGAAGAACTTAAA 1016

Qy     345 AGCCTTTATAATACTGTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACAC 404
           | |||||||| ||       | |||  ||  | ||   | |       |  | ||| ||
Db    1017 TCATTATATAATACAGTAGCAACCCTCTATTGTGTGCATCAAAAGATAGAGATAAAAGAC 1076

Qy     405 ACTGAGGAAGCAAAACAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAA 464
           ||  |||||||   | ||| | |  ||  | | | |          |    || | || |
Db    1077 ACCAAGGAAGCTTTAGAGAAAATAGAGGAAGAGCAAAACAAAAGTAAGAAAAAAGCACAG 1136

Qy     465 ACTATGCCAAAAACAAGTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCC- 523
                 |      |   ||   | || |      |||       |   |||||||||
Db    1137 CAAGCAGTAGCTGACACAGGAAACAGAGGAAACAGCAGCCAAGTCAGCCAAAATTACCCC 1196

Qy     524 --AGTACAACAAATAGGTGGTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAAT 581
             ||| ||  | ||    ||  |    || || | | |  ||   || ||||| ||||||
Db    1197 ATAGTGCAGAACATCCAGGGGCAAATGGTACATCAGGCCATATCACCTAGAACTTTAAAT 1256

Qy     582 GCCTGGGTAAAATTGATAGAGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAG 641
           || |||||||||  |  |||| || |||   ||  |  |||||||| | ||    |||

Db    1257 GCATGGGTAAAAGTAGTAGAAGAGAAGGCTTTCAGCCCAGAAGTAATACCCATGTTTTCA 1316

Qy     642 GCACTGTCAGAAGGTTGCACCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGAC 701
           ||| | ||||||||   ||||||  | ||  | ||    ||| ||||    ||||| |
Db    1317 GCATTATCAGAAGGAGCCACCCCACAAGATTTAAACACCATGCTAAACACAGTGGGGGGA 1376

Qy     702 CATCAAGCGGCTATGCAGATTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGAC 761
           |||||||| || ||||| ||  | | ||| |  || || ||||| |||||||| |||||
Db    1377 CATCAAGCAGCCATGCAAATGTTAAAAGAGACCATCAATGAGGAAGCTGCAGAATGGGAT 1436

Qy     762 TTGCAGCACCCACAACCAG------CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGA 815
           |||  |||   |  |          ||    || ||| ||  | || || ||    |||
Db    1437 AGATTGCATCCAGTGCATGCAGGGCCTATTGCACCAGGCCAGATGAGAGAACCAAGGGGA 1496

Qy     816 TCAGATATTGCAGGAACAACTAGTTCAGTAGATGAACAAATCCAGTGGATGTACAGACAA 875
             || || |||||||| |||||| |  |  | ||||||||    ||||||   | | |
Db    1497 AGTGACATAGCAGGAACTACTAGTACCCTTCAGGAACAAATAGGATGGATG---ACAAAT 1553

Qy     876 CAGAACCCCATACCAGTAGGCAACATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAA 935
           |    || || ||||||||  | || || |  ||||||||    ||||| ||  | |||
Db    1554 AATCCACCTATCCCAGTAGGAGAAATCTATAAAAGATGGATAATCCTGGGATTAAATAAA 1613

Qy     936 TGTGTCAGAATGTATAACCCAACAAACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCA 995
             || || ||||||| |||  | | |||||| ||  ||| |||||| ||||| || ||
Db    1614 ATAGTAAGGATGTATAGCCCTTCCAGCATTCTGGACATAAGACAAGGACCAAAGGAACCC 1673

Qy     996 TTTCAGAGCTATGTAGACAGGTTCTACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCA 1055
           |||     |||||||||| ||||||| |||| | |||||||  || ||  |       |
Db    1674 TTTAGAGACTATGTAGACCGGTTCTATAAAACTCTAAGAGCCGAGCAAGCTTCACAGGAG 1733

Qy    1056 GTAAAGAATTGGATGACTCAAACACTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTA 1115
           ||||| ||||||||||||  ||||  || || | |||||||| ||||||||| |||
Db    1734 GTAAAAAATTGGATGACAGAAACCTTGTTGGTCCAAAATGCGAACCCAGATTGTAAGACT 1793

Qy    1116 GTGCTGAAGGGGCTGGGTGTGAATCCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGA 1175
           |  | || |   ||||         | || |||||||||||| |||| || ||||| |||
Db    1794 ATTTTAAAAGCATTGGGACCAGCAGCTACACTAGAAGAAATGATGACAGCATGTCAGGGA 1853

Qy    1176 GTAGGGGGGCCGGGACAGAAGGCTAGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCA 1235
           || || || || || || || || ||| |  |||| |||||  |||   | |
Db    1854 GTGGGAGGACCTGGTCATAAAGCAAGAGTTTTGGCGGAAGCGATGAGCCAAGTAACAAAT 1913

Qy    1236 CCAGTGCCAATCCCTTTTGCAGCAGCCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGT 1295
           |||   | ||     || | |   ||| |  ||||||    || |||||
Db    1914 TCAGCTACCATAATGATGCAGAGAGGCAATTTTAGGAATCAAAGAAAGATTATCAAGTGC 1973

Qy    1296 TGGAATTGTGGGAAAGAGGGACACTCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGA 1355
           |  |||||||| ||||| || |||   || |   | |||||  |||||  || | |  |||
Db    1974 TTCAATTGTGGCAAAGAAGGGCACATAGCCAAAAATTGCAGGGCCCCTAGGAAAAGGGGC 2033

Qy    1356 TGCTGGAAATGTGGAAAAATGGACCATGTTATGGCCAAATGCCCAGACAGACAGGCGGGT 1415
           || ||||||||||||||    | ||   |||   | ||  | ||  |||||||    |
Db    2034 TGTTGGAAATGTGGAAAGGAAGGACACCAAATGAAAGATTGTACTGAGAGACAGGCTAAT 2093

Qy    1416 TTTTTAGGCCTTGGTCCATGGGGAAAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCAT 1475
           |||||||                |||   |  |       |    ||   |
Db    2094 TTTTTAGGGAAGATCTGGCCTTCCTGCAAGGGAAGGCAGGGAATTTTCCTCAGAGCAGAA 2153

Qy    1476 CAGGGGCTGATGCCAACTGCTCCCCCAGAGGACCCAGCTGTGGATCTGCTAAAGAACTAC 1535
           |||        |||||| || || ||||| ||          |  |  |  | |||
Db    2154 CAG-------AGCCAACAGCCCCACCAGAAGAGAGCTTCAGGTTTGGGGAAGAGACAACA 2206

Qy    1536 ATGCAGTTGGGCAAGCAGCAGAGAGAAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAG 1595
           |  |  |     |||||| |||   |  || |||    |  ||
Db    2207 ACTCCCTATCAGAAGCAGGAGAAGAAGCAGGAGACGATAGACA---------------- 2249

Qy    1596 GAGGTGACAGAGGATTTGCTGCACCTCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTG 1655
           |||       |  |||       |||||| || ||||||||    ||||  | |||||
Db    2250 -AGGACCTGTATCCTTTAGCTTCCCTCAAATCACTCTTTGGCAACGACCCATTGTCACAA 2308

Qy    1656 CTCATATTGAAGGACAGCCTGTAGAAGTATTACTGGATACAGGGGCTGATGATTCTATTG 1715
             || |  || || |     ||||   || | |||||||| || ||||||  |  |
Db    2309 TAAAGATAGGGGGGCAACTAAAGGAAGCTCTATTAGATACAGGAGCAGATGATACAGTAT 2368
```

```
Qy   1716 TAACAGGAATAGAGTTAGGTCCACATTATACCCCAAAAATAGTAGGAGGAATAGGAGGTT 1775
          ||  || |||  | ||      |   |  | ||||||||  |||| ||||| |||||||
Db   2369 TAGAAGAAATGAATTTGCCAGGAAGATGGAAACCAAAAATGATAGGGGGAATTGGAGGTT 2428

Qy   1776 TTATTAATACTAAAGAATACAAAAATGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAG 1835
          |||| ||    | | | ||  |  |  ||   |||||| | |  ||  | |    || ||
Db   2429 TTATCAAAGTAAGACAGTATGATCAGATAACCATAGAAATCTGTGGACATAAAGCTATAG 2488

Qy   1836 GGACAATCATGACAGGGGACACCCCGATTAACATTTTTGGTAGAAATTTGCTAACAGCTC 1895
          | ||| |  |   |||    || ||  | |||||  |||| |||||| || | ||    |
Db   2489 GTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAATCTGTTGACTCAGC 2548

Qy   1896 TGGGGATGTCTCTAAATT 1913
          | |||    || ||||||
Db   2549 TTGGGTGCACTTTAAATT 2566


RESULT 14
US-08-944-449-7
; Sequence 7, Application US/08944449
; Patent No. 5985613
; GENERAL INFORMATION:
;  APPLICANT: KURTH, REINHARD
;  APPLICANT: BAIER, MICHAEL
;  APPLICANT: METZNER, KARIN
;  APPLICANT: WERNER, ALBRECHT
;  TITLE OF INVENTION: Use of an "immunodeficiency-virus suppressing
;  TITLE OF INVENTION: lymphokine (ISL)" to inhibit the replication of
;  TITLE OF INVENTION: viruses, particularly of retroviruses
;  FILE REFERENCE: 8341-7065
;  CURRENT APPLICATION NUMBER: US/08/944,449
;  CURRENT FILING DATE: 1997-10-06
;  EARLIER APPLICATION NUMBER: EP 95113013.2
;  EARLIER FILING DATE: 1995-08-18
;  EARLIER APPLICATION NUMBER: DE 195 13 152.5
;  EARLIER FILING DATE: 1995-04-07
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
;   LENGTH: 9737
;   TYPE: DNA
;   ORGANISM: Human immunodeficiency virus type 1
US-08-944-449-7

  Query Match             26.0%;  Score 497;  DB 2;  Length 9737;
  Best Local Similarity   56.5%;  Pred. No. 1.3e-135;
  Matches 1032;  Conservative    0;  Mismatches  750;  Indels   45;  Gaps    4;

Qy     93 AGGGGTAATAAGATAGAGTGGGAGATGGGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAA 152
          || ||     |||  ||| | ||||||||  | ||||   || || ||    |||    ||
Db    767 AGCGGAGGCTAGAAGGAGAGAGAGATGGGTGCGAGAGCGTCGGTATTAAGCGGGGGAGAA 826

Qy    153 GCAGATGAATTAGAAAAAATTAGGCTACGACCCAACGGAAAGAAAAAGTACATGTTGAAG 212
            |||| |||  |||||||||| || || | ||    |||||||||| || | ||| ||
Db    827 TTAGATAAATGGGAAAAAATTCGGTTAAGGCCAGGGGGAAAGAAAAAATATAAGTTAAAA 886

Qy    213 CATGTAGTATGGGCAGCAAATGAATTAGATAGATTTGGATTAGCAGAAAGCCTGTTGGAG 272
          ||| ||||||||||    |  ||  ||||  |||| | | |         ||||||| ||
Db    887 CATATAGTATGGGCAAGCAGGGAGCTAGAACGATTCGCAGTCAATCCTGGCCTGTTAGAA 946

Qy    273 AACAAAGAAGGATGTCAAAAAATACTTTCGGTCTTAGCTCCATTAGTGCCAACAGGCTCA 332
          |    |||||| ||   | ||||| |         ||   ||||   | |  ||||| |||
Db    947 ACATCAGAAGGCTGCAGACAAATATTGGGACAGCTACAGCCATCCCTTCAGACAGGATCA 1006

Qy    333 GAAAATTTAAAAAGCCTTTATAATACTGTCTGCGTCATCTGGTGCATTCACGCAGAAGAG 392
          ||| |  | | |    | ||||||| ||       | |||  ||  | ||   |
Db   1007 GAAGAACTTAGATCATTATATAATACAGTAGCAACCCTCTATTGTGTACATCAAAGGATA 1066

Qy    393 AAAGTGAAACACACTGAGGAAGCAAAACAGATAGTGCAGAGACACCTAGTGGTGGAAACA 452
          | || ||| ||||  |||||||   | |||   |  ||  | | | |          |
Db   1067 GATGTAAAAGACACCAAGGAAGCTTTAGAGAAGATAGAGGAAGAGCAAAACAAAAGTAAG 1126

Qy    453 GGAACAACAGAAACTATGCCAAAAACAAGTAGACCAACAGCACCATCTAGGGGCAGAGGA 512
          ||   || |         ||    ||  | |  ||  |             | |||  |
Db   1127 AAAAAGGCACAGCAAGCAGCAGCTGCAGCTGGCACAGGAAACAGCAGCCAGGTCAGCCAA 1186

Qy    513 GGAAATTACCCAGTACAACAAATAGGTGGTAACTATGTCCACCTGCCATTAAGCCCGAGA 572
          |      ||| ||  |  ||   ||  |    || || | |  ||   || ||||
Db   1187 AATTACCCTATAGTGCAGAACCTACAGGGGCAAATGGTACATCAGGCCATATCACCTAGA 1246

Qy    573 ACATTAAATGCCTGGGTAAAATTGATAGAGGAAAAGAAATTTGGAGCAGAAGTAGTGCCA 632
          || |||||||| |||||||||  |  |||| ||||||   ||  |  |||||||| | |||
Db   1247 ACTTTAAATGCATGGGTAAAAGTAGTAGAAGAAAAGGCTTTCAGCCCAGAAGTAATACCC 1306

Qy    633 GGATTTCAGGCACTGTCAGAAGGTTGCACCCCCTATGACATTAATCAGATGTTAAATTGT 692
            |||   ||| | ||||||||   ||||||  | ||  | ||    ||| ||||
Db   1307 ATGTTTTCAGCATTATCAGAAGGAGCCACCCCACAAGATTTAAACACCATGCTAAACACA 1366

Qy    693 GTGGGAGACCATCAAGCGGCTATGCAGATTATCAGAGATATTATAAACGAGGAGGCTGCA 752
          |||||  |  |||||||| || |||||  ||   | | ||| | ||| || ||||| ||||||
Db   1367 GTGGGGGGACATCAAGCAGCCATGCAAATGTTAAAAGAGACTATCAATGAGGAAGCTGCA 1426

Qy    753 GATTGGGACTTGCAGCACCCACAACCAG------CTCCACAACAAGGACAACTTAGGGAG 806
          || |||||       ||| |||   |  |       ||      || ||| ||| | || |||
Db   1427 GAATGGGATAGAGTGCATCCAGTGCATGCAGGGCCTATTGCACCAGGCCAAATGAGAGAA 1486

Qy    807 CCGTCAGGATCAGATATTGCAGGAACAACTAGTTCAGTAGATGAACAAATCCAGTGGATG 866
          ||    |||   || || |||||||| ||||||  |   |  | ||||||||    ||||||
Db   1487 CCAAGGGGAAGTGACATAGCAGGAACTACTAGTACCCTTCAGGAACAAATAGGATGGATG 1546

Qy    867 TACAGACAACAGAACCCCATACCAGTAGGCAACATTTACAGGAGATGGATCCAACTGGGG 926
          | | |  |    || || ||||||||  | || || |  ||||||||     |||||
Db   1547 ---ACAAATAATCCACCTATCCCAGTAGGAGAAATCTATAAAAGATGGATAATCCTGGGA 1603

Qy    927 TTGCAAAAATGTGTCAGAATGTATAACCCAACAAACATTCTAGATGTAAAACAAGGGCCA 986
          ||  | |||   || |||||||||| ||| || | |||||| ||  ||| |||||| |||
Db   1604 TTAAATAAAATAGTAAGAATGTATAGCCCTACCAGCATTCTGGACATAAGACAAGGACCA 1663

Qy    987 AAAGAGCCATTTCAGAGCTATGTAGACAGGTTCTACAAAAGTTTAAGAGCAGAACAGACA 1046
          || || || |||      |||||||| |||||| |||| | |||||| ||||| |  |
Db   1664 AAGGAACCCTTTAGAGATTATGTAGACCGGTTCTATAAAACTCTAAGAGCCGAACAAGCT 1723

Qy   1047 GATGCAGCAGTAAAGAATTGGATGACTCAAACACTGCTGATTCAAAATGCTAACCCAGAT 1106
            |  ||||| |||||||||| |||| || || | ||||||| |||||||||
Db   1724 TCACAGGATGTAAAAAATTGGATGACAGAAACCTTGTTGGTCCAAAATGCAAACCCAGAT 1783

Qy   1107 TGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATCCCACCCTAGAAGAAATGCTGACGGCT 1166
          || |||    |  | || |  ||||        | || |||||||||||| |||| ||
Db   1784 TGTAAGACTATTTTAAAAGCATTGGGACCAGCAGCTACACTAGAAGAAATGATGACAGCA 1843

Qy   1167 TGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCTAGATTAATGGCAGAAGCCCTGAAAGAG 1226
          ||||| ||||| ||||| || || || || || ||| |  |||| |||||| |||  |  |
Db   1844 TGTCAGGGAGTGGGGGGACCCGGCCATAAAGCAAGAGTTTTGGCTGAAGCCATGAGCCAA 1903

Qy   1227 GCCCTCGCACCAGTGCCAATCCCTTTTGCAGCAGCCCAACAGAGGGGACCAAGAAAGCCA 1286
          |     ||||     ||      |  || | |  '|||   | |||||| |
Db   1904 GTAACAAATCCAGCTAACATAATGATGCAGAGAGGCAATTTTAGGAACCAAAGAAAGACT 1963

Qy   1287 ATTAAGTGTTGGAATTGTGGGAAAGAGGGACACTCTGCAAGGCAATGCAGAGCCCCAAGA 1346
          |||||||||  |||||||| ||||| || |||    || |    | |||||  |||||| ||
Db   1964 GTTAAGTGTTTCAATTGTGGCAAAGAAGGGCACATAGCCAAAAATTGCAGGGCCCCTAGG 2023

Qy   1347 AGACAGGGATGCTGGAAATGTGGAAAAATGGACCATGTTATGGCCAAATGCCCAGACAGA 1406
          | | |||| || |||| |||||||||     | ||    |||    | ||| | || ||||
Db   2024 AAAAAGGGCTGTTGGAGATGTGGAAGGGAAGGACACCAAATGAAAGATTGCACTGAGAGA 2083

Qy   1407 CAGGCGGGTTTTTTAGGCCTTGGTCCATGGGGAAAGAAGCCCCCGCAATTTCCCCATGGCT 1466
          |||||   |||||||||           | |||  |  |        ||    |  |
Db   2084 CAGGCTAATTTTTTAGGGAAGATCTGGCCTTCCTACAAGGGAAGG------CCAGGGAAT 2137

Qy   1467 CAAGTGCATCAGGGGCTGATGCCAACTGCTCCCCCAGAGGACCCAGCTGTGGATCTGCTA 1526
            | ||      | |    ||||| || || ||||| ||          |  |  |
Db   2138 TTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTCAGGTTTGGGGAG 2197

Qy   1527 AAGAACTACATGCAGTTGGGCAAGCAGCAGAGAGAAAAGCAGAGAGAAAGCAGAGAGAAG 1586
```

```
Db    2198 GAGAAAACAACTCCCTCTCAGAAGC----------------------------AGGAG 2227

Qy    1587 CCTTACAAGGAGGTGACAGAGGATTTGCTGCACCTCAATTCTCTCTTTGGAGGAGACCAG 1646
Db    2228 CCGATAGACAAGGAACTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAACGACCCC 2287

Qy    1647 TAGTCACTGCTCATATTGAAGGACAGCCTGTAGAAGTATTACTGGATACAGGGGCTGATG 1706
Db    2288 TCGTCACAATAAGGATAGGGGGGCAACTAAAGGAAGCTCTATTAGATACAGGAGCAGATG 2347

Qy    1707 ATTCTATTGTAACAGGAATAGAGTTAGGTCCACATTATACCCCAAAAATAGTAGGAGGAA 1766
Db    2348 ATACAGTATTAGAAGAAATGAATTTGCCAGGAAAATGGAAACCAAAAATGATAGGGGGAA 2407

Qy    1767 TAGGAGGTTTTATTAATACTAAAGAATACAAAAATGTAGAAATAGAAGTTTTAGGCAAAA 1826
Db    2408 TTGGAGGTTTTATCAAAGTAAGACAGTACGATCAGATACCTGTAGAAATCTGTGGACATA 2467

Qy    1827 GGATTAAAGGGACAATCATGACAGGGGACACCCCGATTAACATTTTTGGTAGAAATTTGC 1886
Db    2468 AAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAATCTGT 2527

Qy    1887 TAACAGCTCTGGGGATGTCTCTAAATT 1913
Db    2528 TGACTCAGATTGGTTGTACTTTAAATT 2554


RESULT 15
US-09-353-362-7
; Sequence 7, Application US/09353362
; Patent No. 6383739
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Use of an "immunodeficiency-virus suppressing
;    TITLE OF INVENTION:  lymphokine (ISL)" to inhibit the replication of viruses,
;    TITLE OF INVENTION:  in particular of retroviruses
;    NUMBER OF SEQUENCES:  8
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/353,362
;      FILING DATE:  15-JUL-1999
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE 195 13 152.5
;      FILING DATE:  07-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 95113013.2
;      FILING DATE:  18-AUG-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KLESNER, Sharon N.
;      REGISTRATION NUMBER:  36,335
;      REFERENCE/DOCKET NUMBER:  P8341-9012
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 638-5000
;      TELEFAX:  (202) 638-4810
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9737 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-353-362-7

  Query Match             26.0%;  Score 497;  DB 4;  Length 9737;
  Best Local Similarity   56.5%;  Pred. No. 1.3e-135;
  Matches 1032;  Conservative    0;  Mismatches  750;  Indels   45;  Gaps    4;

Qy      93 AGGGGTAATAAGATAGAGTGGGAGATGGGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAA 152
Db     767 AGCGGAGGCTAGAAGGAGAGAGAGATGGGTGCGAGAGCGTCGGTATTAAGCGGGGGAGAA 826

Qy     153 GCAGATGAATTAGAAAAAATTAGGCTACGACCCAACGGAAAGAAAAAGTACATGTTGAAG 212
Db     827 TTAGATAAATGGGAAAAAATTCGGTTAAGGCCAGGGGGAAAGAAAAAATATAAGTTAAAA 886

Qy     213 CATGTAGTATGGGCAGCAAATGAATTAGATAGATTTGGATTAGCAGAAAGCCTGTTGGAG 272
Db     887 CATATAGTATGGGCAAGCAGGGAGCTAGAACGATTCGCAGTCAATCCTGGCCTGTTAGAA 946

Qy     273 AACAAAGAAGGATGTCAAAAAATACTTTCGGTCTTAGCTCCATTAGTGCCAACAGGCTCA 332
Db     947 ACATCAGAAGGCTGCAGACAAATATTGGGACAGCTACAGCCATCCCTTCAGACAGGATCA 1006

Qy     333 GAAAATTTAAAAAGCCTTTATAATACTGTCTGCGTCATCTGGTGCATTCACGCAGAAGAG 392
Db    1007 GAAGAACTTAGATCATTATATAATACAGTAGCAACCCTCTATTGTGTACATCAAAGGATA 1066

Qy     393 AAAGTGAAACACACTGAGGAAGCAAAACAGATAGTGCAGAGACACCTAGTGGTGGAAACA 452
Db    1067 GATGTAAAAGACACCAAGGAAGCTTTAGAGAAGATAGAGGAAGAGCAAAACAAAAGTAAG 1126

Qy     453 GGAACAACAGAAACTATGCCAAAAACAAGTAGACCAACAGCACCATCTAGCGGCAGAGGA 512
Db    1127 AAAAAGGCACAGCAAGCAGCAGCTGCAGCTGGCACAGGAAACAGCAGCCAGGTCAGCCAA 1186

Qy     513 GGAAATTACCCAGTACAACAAATAGGTGGTAACTATGTCCACCTGCCATTAAGCCCGAGA 572
Db    1187 AATTACCCTATAGTGCAGAACCTACAGGGGCAAATGGTACATCAGGCCATATCACCTAGA 1246

Qy     573 ACATTAAATGCCTGGGTAAAATTGATAGAGGAAAAGAAATTTGGAGCAGAAGTAGTGCCA 632
Db    1247 ACTTTAAATGCATGGGTAAAAGTAGTAGAAGAAAAGGCTTTCAGCCCAGAAGTAATACCC 1306

Qy     633 GGATTTCAGGCACTGTCAGAAGGTTGCACCCCCTATGACATTAATCAGATGTTAAATTGT 692
Db    1307 ATGTTTTCAGCATTATCAGAAGGAGCCACCCCACAAGATTTAAACACCATGCTAAACACA 1366

Qy     693 GTGGGAGACCATCAAGCGGCTATGCAGATTATCAGAGATATTATAAACGAGGAGGCTGCA 752
Db    1367 GTGGGGGGACATCAAGCAGCCATGCAAATGTTAAAAGAGACTATCAATGAGGAAGCTGCA 1426

Qy     753 GATTGGGACTTGCAGCACCCACAACCAG------CTCCACAACAAGGACAACTTAGGGAG 806
Db    1427 GAATGGGATAGAGTGCATCCAGTGCATGCAGGGCCTATTGCACCAGGCCAAATGAGAGAA 1486

Qy     807 CCGTCAGGATCAGATATTGCAGGAACAACTAGTTCAGTAGATGAACAAATCCAGTGGATG 866
Db    1487 CCAAGGGGAAGTGACATAGCAGGAACTACTAGTACCCTTCAGGAACAAATAGGATGGATG 1546

Qy     867 TACAGACAACAGAACCCCATACCAGTAGGCAACATTTACAGGAGATGGATCCAACTGGGG 926
Db    1547 ---ACAAATAATCCACCTATCCCAGTAGGAGAAATCTATAAAAGATGGATAATCCTGGGA 1603

Qy     927 TTGCAAAAATGTGTCAGAATGTATAACCCAACAAACATTCTAGATGTAAAACAAGGGCCA 986
Db    1604 TTAAATAAAATAGTAAGAATGTATAGCCCTACCAGCATTCTGGACATAAGACAAGGACCA 1663

Qy     987 AAAGAGCCATTTCAGAGCTATGTAGACAGGTTCTACAAAAGTTTAAGAGCAGAACAGACA 1046
Db    1664 AAGGAACCCTTTAGAGATTATGTAGACCGGTTCTATAAAACTCTAAGAGCCGAACAAGCT 1723

Qy    1047 GATGCAGCAGTAAAGAATTGGATGACTCAAACACTGCTGATTCAAAATGCTAACCCAGAT 1106
Db    1724 TCACAGGATGTAAAAAATTGGATGACAGAAACCTTGTTGGTCCAAAATGCAAACCCAGAT 1783

Qy    1107 TGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATCCCACCCTAGAAGAAATGCTGACGGCT 1166
Db    1784 TGTAAGACTATTTTAAAAGCATTGGGACCAGCAGCTACACTAGAAGAAATGATGACAGCA 1843
```

```
Qy   1167 TGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCTAGATTAATGGCAGAAGCCCTGAAAGAG 1226
          ||||| ||||| ||||| || || || || || ||| |  |||| ||||||| |||   |
Db   1844 TGTCAGGGAGTGGGGGGACCCGGCCATAAAGCAAGAGTTTTGGCTGAAGCCATGAGCCAA 1903

Qy   1227 GCCCTCGCACCAGTGCCAATCCCTTTTGCAGCAGCCCAACAGAGGGGACCAAGAAAGCCA 1286
          |        ||||     ||      |       || | |    |||   | |||||||  |
Db   1904 GTAACAAATCCAGCTAACATAATGATGCAGAGAGGCAATTTTAGGAACCAAAGAAAGACT 1963

Qy   1287 ATTAAGTGTTGGAATTGTGGGAAAGAGGGACACTCTGCAAGGCAATGCAGAGCCCCAAGA 1346
           |||||||||  ||||||||  |||||  || |||   || |   | |||||  |||||  ||
Db   1964 GTTAAGTGTTTCAATTGTGGCAAAGAAGGGCACATAGCCAAAAATTGCAGGGCCCCTAGG 2023

Qy   1347 AGACAGGGATGCTGGAAATGTGGAAAAATGGACCATGTTATGGCCAAATGCCCAGACAGA 1406
          | | |||| || |||| ||||||||      |  ||    |||    | ||| | || |||
Db   2024 AAAAAGGGCTGTTGGAGATGTGGAAGGGAAGGACACCAAATGAAAGATTGCACTGAGAGA 2083

Qy   1407 CAGGCGGGTTTTTTAGGCCTTGGTCCATGGGGAAAGAAGCCCCGCAATTTCCCCATGGCT 1466
          |||||   |||||||||                | |||   |       ||   |  |
Db   2084 CAGGCTAATTTTTTAGGGAAGATCTGGCCTTCCTACAAGGGAAGG------CCAGGGAAT 2137

Qy   1467 CAAGTGCATCAGGGGCTGATGCCAACTGCTCCCCCAGAGGACCCAGCTGTGGATCTGCTA 1526
              | ||     | |    |||||| || || ||||| ||         |  |  |
Db   2138 TTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTCAGGTTTGGGGAG 2197

Qy   1527 AAGAACTACATGCAGTTGGGCAAGCAGCAGAGAGAAAAGCAGAGAGAAAGCAGAGAGAAG 1586
           ||||    |  | |     ||||                                ||  ||
Db   2198 GAGAAAACAACTCCCTCTCAGAAGC----------------------------AGGAG 2227

Qy   1587 CCTTACAAGGAGGTGACAGAGGATTTGCTGCACCTCAATTCTCTCTTTGGAGGAGACCAG 1646
          ||     |  |||       |   |||      |||||  || ||||||||    ||||
Db   2228 CCGATAGACAAGGAACTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAACGACCCC 2287

Qy   1647 TAGTCACTGCTCATATTGAAGGACAGCCTGTAGAAGTATTACTGGATACAGGGGCTGATG 1706
          | |||||       || |  || || |     ||||   || | |||||||| || ||||
Db   2288 TCGTCACAATAAGGATAGGGGGGCAACTAAAGGAAGCTCTATTAGATACAGGAGCAGATG 2347

Qy   1707 ATTCTATTGTAACAGGAATAGAGTTAGGTCCACATTATACCCCAAAAATAGTAGGAGGAA 1766
          || |  |  ||  || |||  | ||      | | |  |  ||||||||  |||| ||||
Db   2348 ATACAGTATTAGAAGAAATGAATTTGCCAGGAAAATGGAAACCAAAAATGATAGGGGGAA 2407

Qy   1767 TAGGAGGTTTTATTAATACTAAAGAATACAAAAATGTAGAAATAGAAGTTTTAGGCAAAA 1826
          | ||||||||||| ||    | | | ||| |  |  ||    ||||| | |  ||  | |
Db   2408 TTGGAGGTTTTATCAAAGTAAGACAGTACGATCAGATACCTGTAGAAATCTGTGGACATA 2467

Qy   1827 GGATTAAAGGGACAATCATGACAGGGGACACCCCGATTAACATTTTTGGTAGAAATTTGC 1886
              || ||| ||| |  |   |||    || ||  | |||||  |||| ||||||  ||
Db   2468 AAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAATCTGT 2527

Qy   1887 TAACAGCTCTGGGGATGTCTCTAAATT 1913
          | ||     | ||     || ||||||
Db   2528 TGACTCAGATTGGTTGTACTTTAAATT 2554
```

Search completed: November 16, 2002, 07:54:10
Job time : 116.345 secs

m33262_copy_1193_3105.rnpb

Mon Nov 18 08:37:02 2002

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 16, 2002, 07:33:33 ; Search time 48.6447 Seconds
(without alignments)
14531.422 Million cell updates/sec

Title: M33262_COPY_1193_3105
Perfect score: 1913
Sequence: 1 GGTACCAGACGGCGTGAGGA.....TCTGGGGATGTCTCTAAATT 1913

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 328367 seqs, 184756068 residues

Total number of hits satisfying chosen parameters: 656734

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1490.2	77.9	1532	10	US-09-872-733-5	Sequence 5, Appli
2	1411.4	73.8	1533	10	US-09-872-733-4	Sequence 4, Appli
3	509.6	26.6	8933	10	US-09-943-286-3	Sequence 3, Appli
4	509.6	26.6	8933	10	US-09-943-286-4	Sequence 4, Appli
5	509.6	26.6	8933	10	US-09-943-286-9	Sequence 9, Appli
6	500	26.1	9544	10	US-09-798-675-4	Sequence 4, Appli
7	500	26.1	9918	10	US-09-798-675-5	Sequence 5, Appli
8	498.2	26.0	1800	9	US-10-003-035-58	Sequence 58, Appl
9	485.4	25.4	9719	10	US-09-737-190A-1	Sequence 1, Appli
10	481.4	25.2	4307	10	US-09-999-183-1	Sequence 1, Appli
11	455	23.8	8366	10	US-09-872-733-6	Sequence 6, Appli
12	450.4	23.5	4516	10	US-09-968-355-24	Sequence 24, Appl
13	437.2	22.9	4438	10	US-09-968-355-21	Sequence 21, Appl
14	437.2	22.9	4594	10	US-09-968-355-15	Sequence 15, Appl
15	437	22.8	1503	10	US-09-968-355-25	Sequence 25, Appl
16	437	22.8	1596	10	US-09-968-355-22	Sequence 22, Appl
17	437	22.8	1752	10	US-09-968-355-16	Sequence 16, Appl
18	434.8	22.7	4390	10	US-09-968-355-18	Sequence 18, Appl
19	434.6	22.7	1548	10	US-09-968-355-19	Sequence 19, Appl
20	431.2	22.5	9793	9	US-09-886-156-56	Sequence 56, Appl
21	423.6	22.1	4338	10	US-09-872-733-1	Sequence 1, Appli
22	400.4	20.9	12523	10	US-09-991-258-1	Sequence 1, Appli
23	398.4	20.8	1476	10	US-09-991-258-4	Sequence 4, Appli
24	373.4	19.5	1733	9	US-09-886-156-58	Sequence 58, Appl
25	371.8	19.4	1496	9	US-10-003-035-17	Sequence 17, Appl
26	368.8	19.3	1733	9	US-09-886-156-57	Sequence 57, Appl
27	364.4	19.0	1092	9	US-10-003-035-34	Sequence 34, Appl
28	361.4	18.9	1179	9	US-10-003-035-35	Sequence 35, Appl
29	361.4	18.9	1308	9	US-10-003-035-36	Sequence 36, Appl
30	356.4	18.6	1532	10	US-09-818-443-1	Sequence 1, Appli
31	356.4	18.6	4307	10	US-09-999-183-2	Sequence 2, Appli
32	323.6	16.9	1482	10	US-09-818-443-4	Sequence 4, Appli
33	311.4	16.3	845	10	US-09-756-551A-3	Sequence 3, Appli
34	278.6	14.6	699	9	US-10-003-035-46	Sequence 46, Appl
35	278.6	14.6	786	9	US-10-003-035-47	Sequence 47, Appl
36	278.6	14.6	915	9	US-10-003-035-48	Sequence 48, Appl
37	162.6	8.5	2601	10	US-09-735-487-7	Sequence 7, Appli
38	162.6	8.5	2601	10	US-09-735-487-9	Sequence 9, Appli
39	162.6	8.5	2601	10	US-09-735-487-11	Sequence 11, Appl
40	162.6	8.5	2601	10	US-09-735-487-13	Sequence 13, Appl
41	106.2	5.6	559	9	US-10-035-098-31	Sequence 31, Appl
c 42	106.2	5.6	559	9	US-10-035-098-33	Sequence 33, Appl
43	104	5.4	8937	10	US-09-872-733-8	Sequence 8, Appli
44	103.2	5.4	399	9	US-10-003-035-40	Sequence 40, Appl
45	100.2	5.2	486	9	US-10-003-035-41	Sequence 41, Appl

ALIGNMENTS

RESULT 1
US-09-872-733-5
; Sequence 5, Application US/09872733
; Patent No. US20010036655A1
; GENERAL INFORMATION:
; APPLICANT: The Government of the United States of America, as
; TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, SIV GAG AND
; TITLE OF INVENTION: SIV ENV GENES
; FILE REFERENCE: 2026-4287US1 HIV GAG/POL,SIV GAG & ENV
; CURRENT APPLICATION NUMBER: US/09/872,733
; CURRENT FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: PCT/US00/34985
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/173,036
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1532
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Consensus
; OTHER INFORMATION: sequence of mutated Simian Immunodeficiency Virus
; OTHER INFORMATION: Gag gene (SIVgagDX) with wild-type SIV 239 Gag
; OTHER INFORMATION: gene
US-09-872-733-5

Query Match 77.9%; Score 1490.2; DB 10; Length 1532;
Best Local Similarity 94.9%; Pred. No. 0;
Matches 1455; Conservative 77; Mismatches 0; Indels 1; Gaps 1;

Qy 117 ATGGGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGG 176
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 ATGGGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGG 60

Qy 177 CTACGACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAA 236
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 61 CTACGACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAA 120

Qy 237 TTAGATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATA 296
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
Db    121 TTAGATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATA 180

Qy    297 CTTTCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAAT 356
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CTTTCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAAT 240

Qy    357 ACTGTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCA 416
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ACTGTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCA 300

Qy    417 AAACAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAA 476
          ||||||||||||||||||||||||||||||||||||||||:||:|||||:|||||:||:
Db    301 AAACAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACMACMGAAACYATGCCRAAR 360

Qy    477 ACAAGTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATA 536
          ||:::|:|||||||||||||||||||||||||||||||||||:||||||||||:||::||:
Db    361 ACMWSTMGACCAACAGCACCATCTAGCGGCAGAGGAGGAAAYTACCCAGTACARCARATM 420

Qy    537 GGTGGTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTG 596
          ||||||||||:|||||||||||:|:::||||||||::|:||:||:|||||:||::||
Db    421 GGTGGTAACTAYGTCCACCTGCCAYTRWSCCCGAGAACMYTRAAYGCYTGGGTMAARYTG 480

Qy    597 ATAGAGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGT 656
          ||:|||||:|||||:||:|||||||||||||||||||||:|||||||||||||||||||
Db    481 ATMGAGGARAAGAARTTYGGAGCAGAAGTAGTGCCAGGATTYCAGGCACTGTCAGAAGGT 540

Qy    657 TGCACCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATG 716
          ||||||||||:|||||:||:|||||:|:||:||:||:|||||||||||:||||||||||
Db    541 TGCACCCCCTAYGACATYAAYCAGATGYTRAAYTGYGTKGGAGACCATCARGCGGCTATG 600

Qy    717 CAGATTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAA 776
          |||||:|||:|:||:||:||:|||||||||||||||| ||||||||||||||||||||||
Db    601 CAGATYATCMGWGAYATYATMAACGAGGAGGCTGCAGA-TGGGACTTGCAGCACCCACAA 659

Qy    777 CCAGCTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACT 836
          ||||||||||||||||||||||||||||||||||||||||||||:||:|||||||||:||:
Db    660 CCAGCTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGAYATYGCAGGAACMACY 719

Qy    837 AGTTCAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGC 896
          :::|||||:||:|||||:|||||||||||||||||:|:||:|||||||||:||:|||||||
Db    720 WSYTCAGTWGAYGAACARATCCAGTGGATGTACMGWCARCAGAACCCSATMCCAGTAGGC 779

Qy    897 AACATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCA 956
          |||||:|||:|::|||||||||:|||||::||||:||:||:||::|:|||||:|||||:
Db    780 AACATYTACMGKMGATGGATCCARCTGGGKYTGCARAARTGYGTYMGWATGTAYAACCCR 839

Qy    957 ACAAACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGG 1016
          ||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    840 ACMAACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGG 899

Qy   1017 TTCTACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAA 1076
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    900 TTCTACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAA 959

Qy   1077 ACACTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTG 1136
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    960 ACACTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTG 1019

Qy   1137 AATCCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAG 1196
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1020 AATCCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAG 1079

Qy   1197 GCTAGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCA 1256
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1080 GCTAGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCA 1139

Qy   1257 GCAGCCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGA 1316
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1140 GCAGCCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGA 1199

Qy   1317 CACTCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATG 1376
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1200 CACTCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATG 1259

Qy   1377 GACCATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGG 1436
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1260 GACCATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGG 1319

Qy   1437 GGAAAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGCT 1496
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1320 GGAAAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGCT 1379

Qy   1497 CCCCCAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAG 1556
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1380 CCCCCAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAG 1439

Qy   1557 AGAGAAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTG 1616
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1440 AGAGAAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTG 1499

Qy   1617 CACCTCAATTCTCTCTTTGGAGGAGACCAGTAG 1649
          ||||||||||||||||||||||||||||||||
Db   1500 CACCTCAATTCTCTCTTTGGAGGAGACCAGTAG 1532
```

RESULT 2
US-09-872-733-4
; Sequence 4, Application US/09872733
; Patent No. US20010036655A1
; GENERAL INFORMATION:
; APPLICANT: The Government of the United States of America, as
; TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, SIV GAG AND
; TITLE OF INVENTION: SIV ENV GENES
; FILE REFERENCE: 2026-4287US1 HIV GAG/POL,SIV GAG & ENV
; CURRENT APPLICATION NUMBER: US/09/872,733
; CURRENT FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: PCT/US00/34985
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/173,036
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1533
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Mutated
; OTHER INFORMATION: Simian Immunodeficiency Virus Gag gene
US-09-872-733-4

Query Match 73.8%; Score 1411.4; DB 10; Length 1533;
Best Local Similarity 95.0%; Pred. No. 0;
Matches 1457; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

```
Qy     117 ATGGGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGG 176
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGGGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGG 60

Qy     177 CTACGACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAA 236
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CTACGACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAA 120

Qy     237 TTAGATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATA 296
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TTAGATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATA 180

Qy     297 CTTTCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAAT 356
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CTTTCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAAT 240

Qy     357 ACTGTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCA 416
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 ACTGTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCA 300

Qy     417 AAACAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAA 476
```

```
Db    301 AAACAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACCACCGAAACCATGCCGAAG 360

Qy    477 ACAAGTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATA 536

Db    361 ACCTCTCGACCAACAGCACCATCTAGCGGCAGAGGAGGAAACTACCCAGTACAGCAGATC 420

Qy    537 GGTGGTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTG 596

Db    421 GGTGGTAACTACGTCCACCTGCCACTGTCCCCGAGAACCCTGAACGCTTGGGTCAAGCTG 480

Qy    597 ATAGAGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGT 656

Db    481 ATCGAGGAGAAGAAGTTCGGAGCAGAAGTAGTGCCAGGATTCCAGGCACTGTCAGAAGGT 540

Qy    657 TGCACCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATG 716

Db    541 TGCACCCCCTACGACATCAACCAGATGCTGAACTGCGTTGGAGACCATCAGGCGGCTATG 600

Qy    717 CAGATTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAA 776

Db    601 CAGATCATCCGTGACATCATCAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAA 660

Qy    777 CCAGCTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGAACAACT 836

Db    661 CCAGCTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGACATCGCAGGAACCACC 720

Qy    837 AGTTCAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCAGTAGGC 896

Db    721 TCCTCAGTTGACGAACAGATCCAGTGGATGTACCGTCAGCAGAACCCGATCCCAGTAGGC 780

Qy    897 AACATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTATAACCCA 956

Db    781 AACATCTACCGTCGATGGATCCAGCTGGGTCTGCAGAAGTGCGTCCGTATGTACAACCCG 840

Qy    957 ACAAACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGG 1016

Db    841 ACCAACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGG 900

Qy   1017 TTCTACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAA 1076

Db    901 TTCTACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAA 960

Qy   1077 ACACTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTG 1136

Db    961 ACACTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTG 1020

Qy   1137 AATCCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAG 1196

Db   1021 AATCCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAG 1080

Qy   1197 GCTAGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCA 1256

Db   1081 GCTAGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCTTTTGCA 1140

Qy   1257 GCAGCCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGA 1316

Db   1141 GCAGCCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAAGAGGGA 1200

Qy   1317 CACTCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATG 1376

Db   1201 CACTCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAAATG 1260

Qy   1377 GACCATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGG 1436

Db   1261 GACCATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGTCCATGG 1320

Qy   1437 GGAAAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGCT 1496

Db   1321 GGAAAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCAACTGCT 1380

Qy   1497 CCCCCAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAG 1556
```

```
Db   1381 CCCCCAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAG 1440

Qy   1557 AGAGAAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTG 1616

Db   1441 AGAGAAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTG 1500

Qy   1617 CACCTCAATTCTCTCTTTGGAGGAGACCAGTAG 1649

Db   1501 CACCTCAATTCTCTCTTTGGAGGAGACCAGTAG 1533


RESULT 3
US-09-943-286-3
; Sequence 3, Application US/09943286
; Patent No. US20020106668A1
; GENERAL INFORMATION:
;  APPLICANT: Nunomura, Kiyotada
;  TITLE OF INVENTION: POLYNUCLEOTIDE AMPLIFICATION METHOD
;  FILE REFERENCE: GP104-02.UT
;  CURRENT APPLICATION NUMBER: US/09/943,286
;  CURRENT FILING DATE:  2001-08-30
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
;   LENGTH: 8933
;   TYPE: RNA
;   ORGANISM: Human Immunodeficiency Virus
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)...(8933)
;   OTHER INFORMATION: Sequence of transcripts produced from the BH10
;   OTHER INFORMATION: plasmid.
US-09-943-286-3

  Query Match             26.6%;  Score 509.6;  DB 10;  Length 8933;
  Best Local Similarity   45.2%;  Pred. No. 9.3e-127;
  Matches  821;  Conservative  214;  Mismatches  764;  Indels   17;  Gaps    5;

Qy    105 ATAGAGTGGGAGATGGGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTA 164

Db    101 AGAAGGAGAGAGAUGGGUGCGAGAGCGUCAGUAUUAAGCGGGGGAGAAUUAGAUCGAUGG 160

Qy    165 GAAAAAATTAGGCTACGACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGG 224

Db    161 GAAAAAAUUCGGUUAAGGCCAGGGGGAAAGAAAAAAUAUAAAUUAAAACAUAUAGUAUGG 220

Qy    225 GCAGCAAATGAATTAGATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGA 284

Db    221 GCAAGCAGGGAGCUAGAACGAUUCGCAGUUAAUCCUGGCCUGUUAGAAACAUCAGAAGGC 280

Qy    285 TGTCAAAAAATACTTTCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAA 344

Db    281 UGUAGACAAAUACUGGGACAGCUACAACCAUCCCUUCAGACAGGAUCAGAAGAACUUAGA 340

Qy    345 AGCCTTTATAATACTGTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACAC 404

Db    341 UCAUUAUAUAAUACAGUAGCAACCCUCUAUUGUGUGCAUCAAAGGAUAGAGAUAAAAGAC 400

Qy    405 ACTGAGGAAGCAAAACAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAA 464

Db    401 ACCAAGGAAGC------UUUAGACAAGAUAGAGGAAGAGCAAAACAAAAGUAAGAAAAAA 454

Qy    465 ACTATGCCAAAAACAAGTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCA 524

Db    455 GCACAGCAAGCAGCAGCUGACACAGGACACAGCAGUCAGGUCAGCCAAAAUUACCCUAUA 514

Qy    525 GTACAACAAATAGGTGGTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCC 584

Db    515 GUGCAGAACAUCCAGGGGCAAAUGGUACAUCAGGCCAUAUCACCUAGAACUUUAAAUGCA 574

Qy    585 TGGGTAAAATTGATAGAGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCA 644

Db    575 UGGGUAAAAGUAGUAGAAGAGAAGGCUUUCAGCCCAGAAGUAAUACCCAUGUUUUCAGCA 634
```

```
Qy    645 CTGTCAGAAGGTTGCACCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCAT 704
Db    635 UUAUCAGAAGGAGCCACCCCACAAGAUUUAAACACCAUGCUAAACACAGUGGGGGGACAU 694

Qy    705 CAAGCGGCTATGCAGATTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTG 764
Db    695 CAAGCAGCCAUGCAAAUGUUAAAAGAGACCAUCAAUGAGGAAGCUGCAGAAUGGGAUAGA 754

Qy    765 CAGCACCCACAACCAG------CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCA 818
Db    755 GUACAUCCAGUGCAUGCAGGGCCUAUUGCACCAGGCCAGAUGAGAGAACCAAGGGGAAGU 814

Qy    819 GATATTGCAGGAACAACTAGTTCAGTAGATGAACAAATCCAGTGGATGTACAGACAACAG 878
Db    815 GACAUAGCAGGAACUACUAGUACCCUUCAGGAACAAAUAGGAUGGAUG---ACAAAUAAU 871

Qy    879 AACCCCATACCAGTAGGCAACATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGT 938
Db    872 CCACCUAUCCCAGUAGGAGAAAUUUAUAAAAGAUGGAUAAUCCUGGGAUUAAAUAAAAUA 931

Qy    939 GTCAGAATGTATAACCCAACAAACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTT 998
Db    932 GUAAGAAUGUAUAGCCCUACCAGCAUUCUGGACAUAAGACAAGGACCAAAAGAACCUUUU 991

Qy    999 CAGAGCTATGTAGACAGGTTCTACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTA 1058
Db    992 AGAGACUAUGUAGACCGGUUCUAUAAAACUCUAAGAGCCGAGCAAGCUUCACAGGAGGUA 1051

Qy   1059 AAGAATTGGATGACTCAAACACTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTG 1118
Db   1052 AAAAAUUGGAUGACAGAAACCUUGUUGGUCCAAAAUGCGAACCCAGAUUGUAAGACUAUU 1111

Qy   1119 CTGAAGGGGCTGGGTGTGAATCCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTA 1178
Db   1112 UUAAAAGCAUUGGGACCAGCGGCUACACUAGAAGAAAUGAUGACAGCAUGUCAGGGAGUA 1171

Qy   1179 GGGGGGCCGGGACAGAAGGCTAGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCA 1238
Db   1172 GGAGGACCCGGCCAUAAGGCAAGAGUUUUGGCUGAAGCAAUGAGCCAAGUAACAAAUACA 1231

Qy   1239 GTGCCAATCCCTTTTGCAGCAGCCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGG 1298
Db   1232 GCUACCAUAAUGAUGCAGAGAGGCAAUUUUAGGAACCAAAGAAAGAUGGUUAAGUGUUUC 1291

Qy   1299 AATTGTGGGAAAGAGGGACACTCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGC 1358
Db   1292 AAUUGUGGCAAAGAAGGGCACACAGCCAGAAAUUGCAGGGCCCCUAGGAAAAAGGGCUGU 1351

Qy   1359 TGGAAATGTGGAAAAATGGACCATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTT 1418
Db   1352 UGGAAAUGUGGAAAGGAAGGACACCAAAUGAAAGAUUGUACUGAGAGACAGGCUAAUUUU 1411

Qy   1419 TTAGGCCTTGGTCCATGGGGAAAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAG 1478
Db   1412 UUAGGGAAGAUCUGGCCUUCCUACAAGGGAAGGCCAGGGAAUUUUCUUCAGAGCAGACCA 1471

Qy   1479 GGGCTGATGCCAACTGCTCCCCCAGAGGACCCAGCTGTGG-ATCTGCTAAAGAACTACAT 1537
Db   1472 GAGCCAACAGCCCCACCAUUUCUUCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAG 1531

Qy   1538 GCAGTTGGGCAAGCAGCAGAGAGAAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGA 1597
Db   1532 AGCUUCAGGUCUGGGGUAGAGACAACAACUCCCCCUCAGAAG-CAGGAGCCGAUAGACAA 1590

Qy   1598 GGTGACAGAGGATTTGCTGCACCTCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCT 1657
Db   1591 GGAACUGUAUCCUUUAACUUCCCUCAGAUCACUCUUUGGCAACGACCCCUCGUCACAAUA 1650

Qy   1658 CATATTGAAGGACAGCCTGTAGAAGTATTACTGGATACAGGGGCTGATGATTCTATTGTA 1717
Db   1651 AAGAUAGGGGGGCAACUAAAGGAAGCUCUAUUAGAUACAGGAGCAGAUGAUACAGUAUUA 1710

Qy   1718 ACAGGAATAGAGTTAGGTCCACATTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTT 1777
Db   1711 GAAGAAAUGAGUUUGCCAGGAAGAUGGAAACCAAAAAUGAUAGGGGGAAUUGGAGGUUUU 1770

Qy   1778 ATTAATACTAAAGAATACAAAAATGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGG 1837
Db   1771 AUCAAAGUAAGACAGUAUGAUCAGAUACUCAUAGAAAUCUGUGGACAUAAAGCUAUAGGU 1830

Qy   1838 ACAATCATGACAGGGGACACCCCGATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTG 1897
Db   1831 ACAGUAUUAGUAGGACCUACACCUGUCAACAUAAUUGGAAGAAAUCUGUUGACUCAGAUU 1890

Qy   1898 GGGATGTCTCTAAATT 1913
Db   1891 GGUUGCACUUUAAAUU 1906


RESULT 4
US-09-943-286-4
; Sequence 4, Application US/09943286
; Patent No. US20020106668A1
; GENERAL INFORMATION:
;  APPLICANT: Nunomura, Kiyotada
;  TITLE OF INVENTION: POLYNUCLEOTIDE AMPLIFICATION METHOD
;  FILE REFERENCE: GP104-02.UT
;  CURRENT APPLICATION NUMBER: US/09/943,286
;  CURRENT FILING DATE:  2001-08-30
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
;   LENGTH: 8933
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Sequence of the IAC-Asrc pseudo target
;   NAME/KEY: mutation
;   LOCATION: (4135)...(4155)
;   OTHER INFORMATION: Mutated positions: 4135, 4140-1, 4145, 4150,
;   OTHER INFORMATION: 4152-3, 4155
US-09-943-286-4

  Query Match           26.6%;  Score 509.6;  DB 10;  Length 8933;
  Best Local Similarity   45.2%;  Pred. No. 9.3e-127;
  Matches  821;  Conservative  214;  Mismatches  764;  Indels   17;  Gaps    5;

Qy    105 ATAGAGTGGGAGATGGGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTA 164
Db    101 AGAAGGAGAGAGAUGGGUGCGAGAGCGUCAGUAUUAAGCGGGGGAGAAUUAGAUCGAUGG 160

Qy    165 GAAAAAATTAGGCTACGACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGG 224
Db    161 GAAAAAAUUCGGUUAAGGCCAGGGGGAAAGAAAAAAUAUAAAUUAAAACAUAUAGUAUGG 220

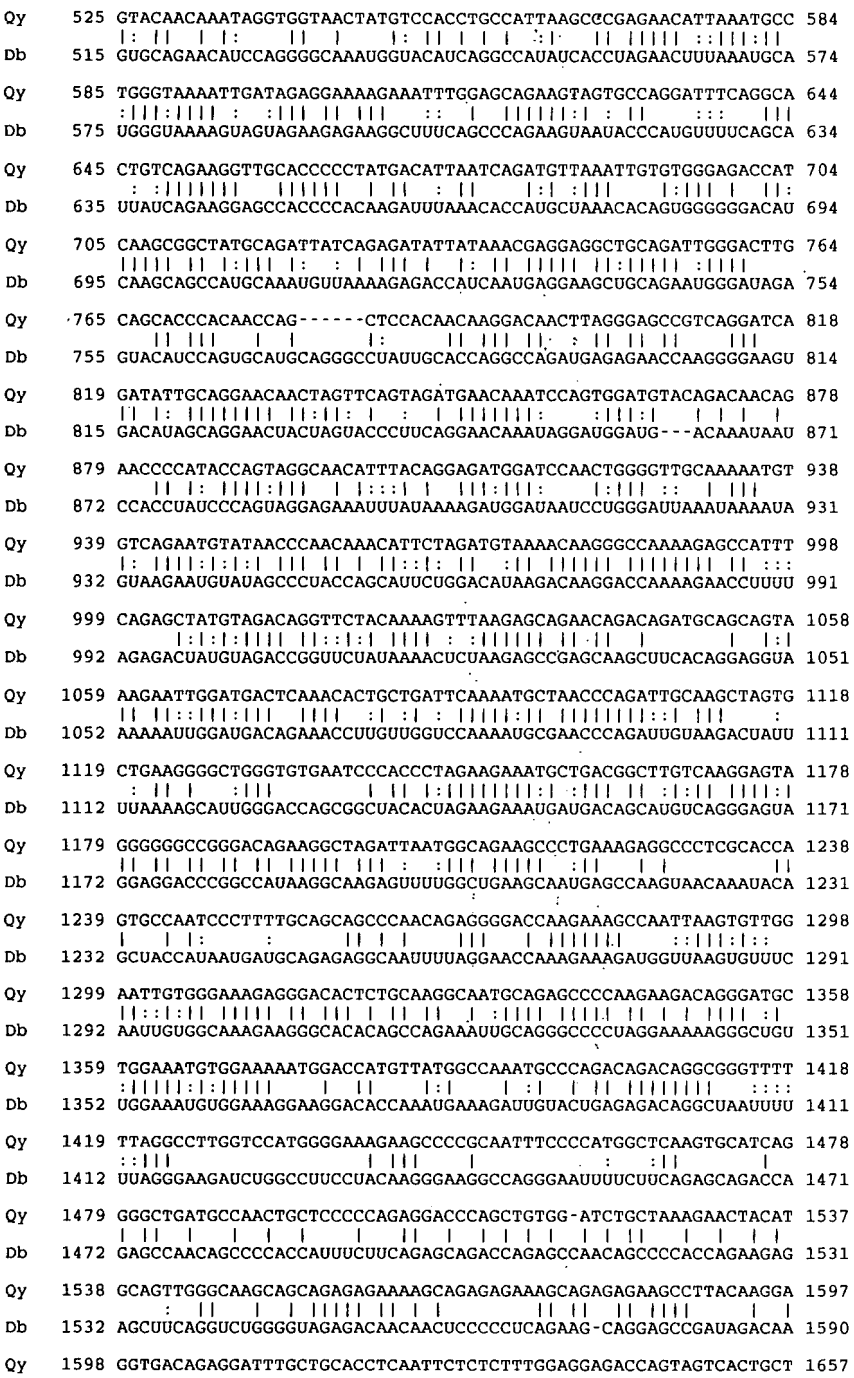
Qy    225 GCAGCAAATGAATTAGATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGA 284
Db    221 GCAAGCAGGGAGCUAGAACGAUUCGCAGUUAAUCCUGGCCUGUUAGAAACAUCAGAAGGC 280

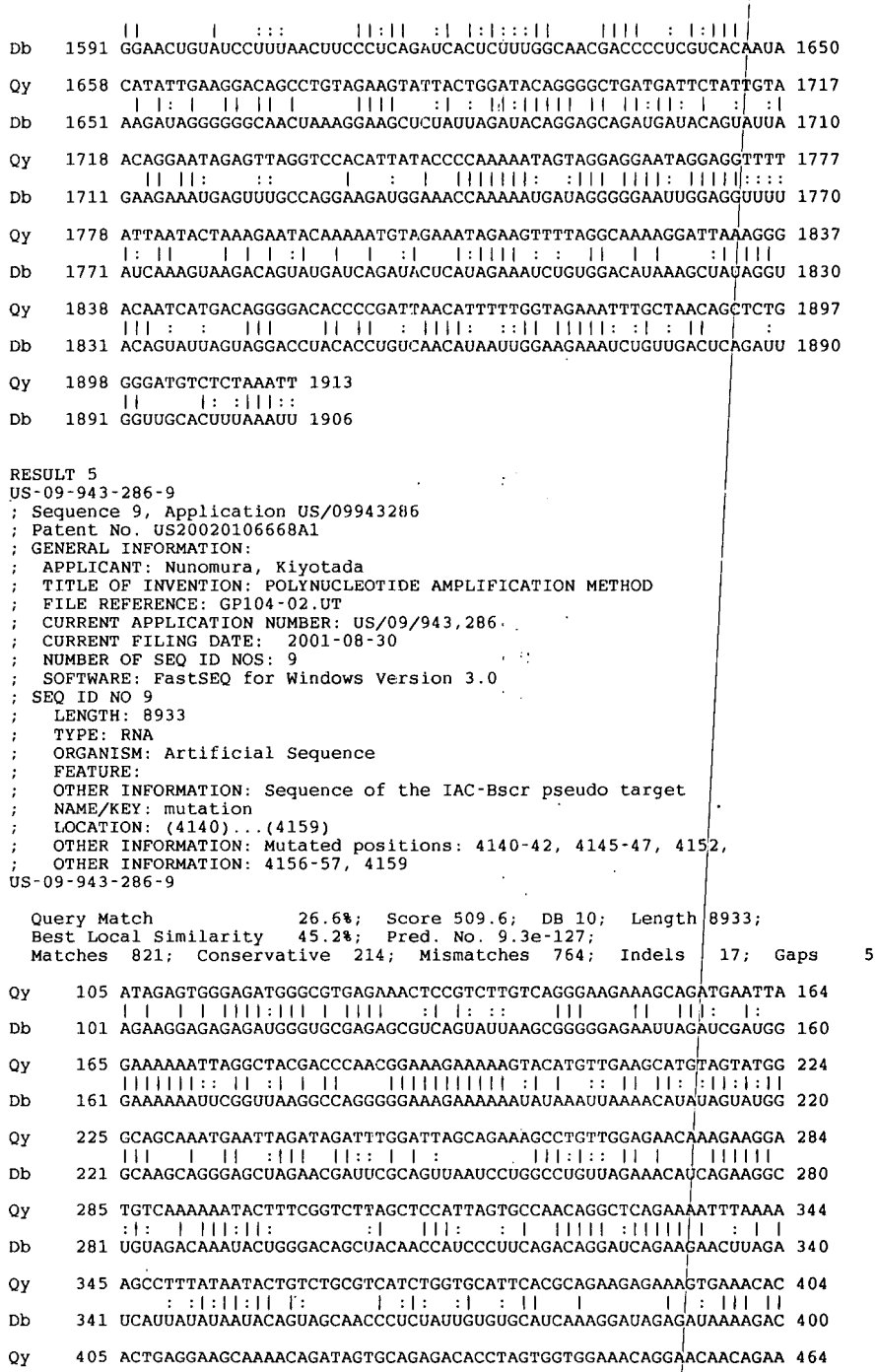
Qy    285 TGTCAAAAAATACTTTCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAA 344
Db    281 UGUAGACAAAUACUGGGACAGCUACAACCAUCCCUUCAGACAGGAUCAGAAGAACUUAGA 340

Qy    345 AGCCTTTATAATACTGTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACAC 404
Db    341 UCAUUAUAUAAUACAGUAGCAACCCUCUAUUGUGUGCAUCAAAGGAUAGAGAUAAAAGAC 400

Qy    405 ACTGAGGAAGCAAAACAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAA 464
Db    401 ACCAAGGAAGC------UUUAGACAAGAUAGAGGAAGAGCAAAACAAAAGUAAGAAAAAA 454

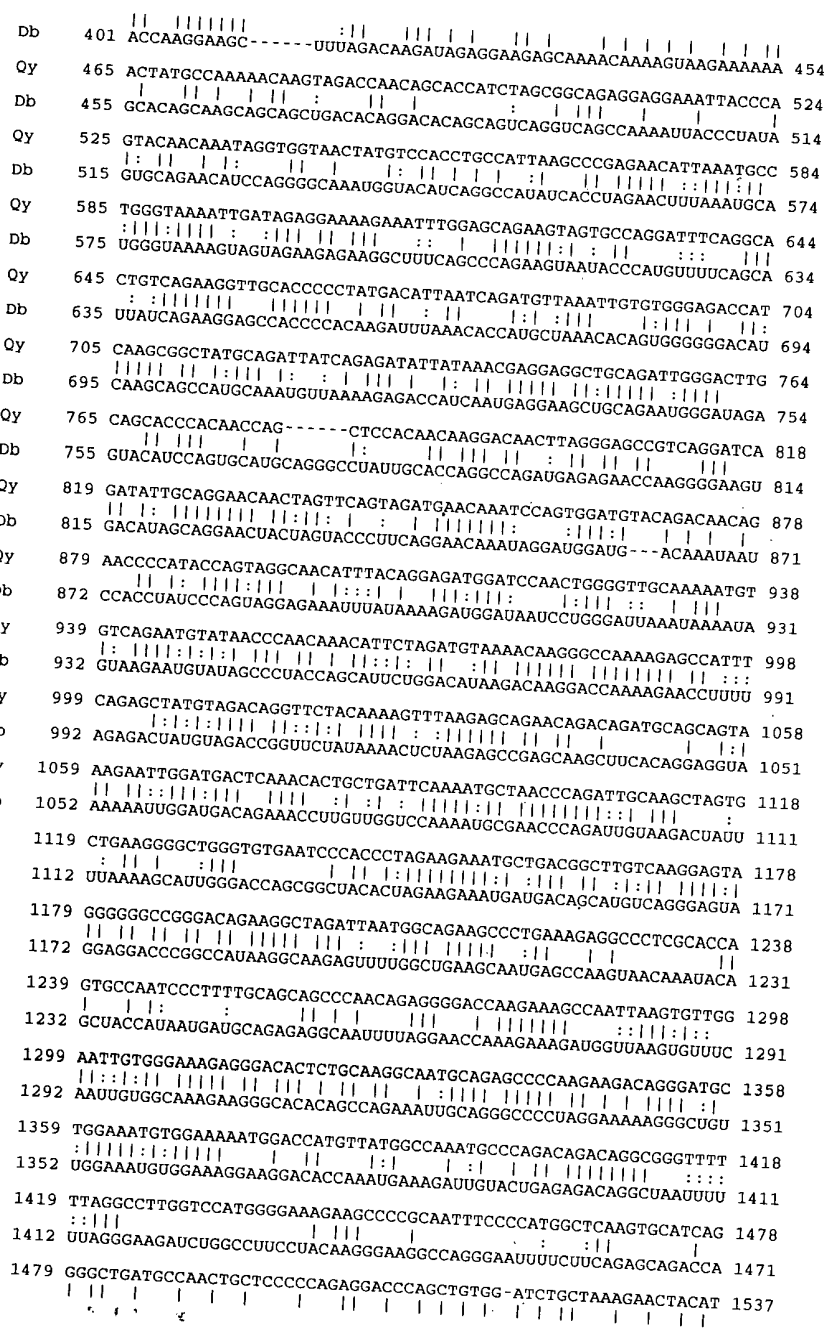
Qy    465 ACTATGCCAAAAACAAGTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCA 524
Db    455 GCACAGCAAGCAGCAGCUGACACAGGACACAGCAGUCAGGUCAGCCAAAAUUACCCUAUA 514
```

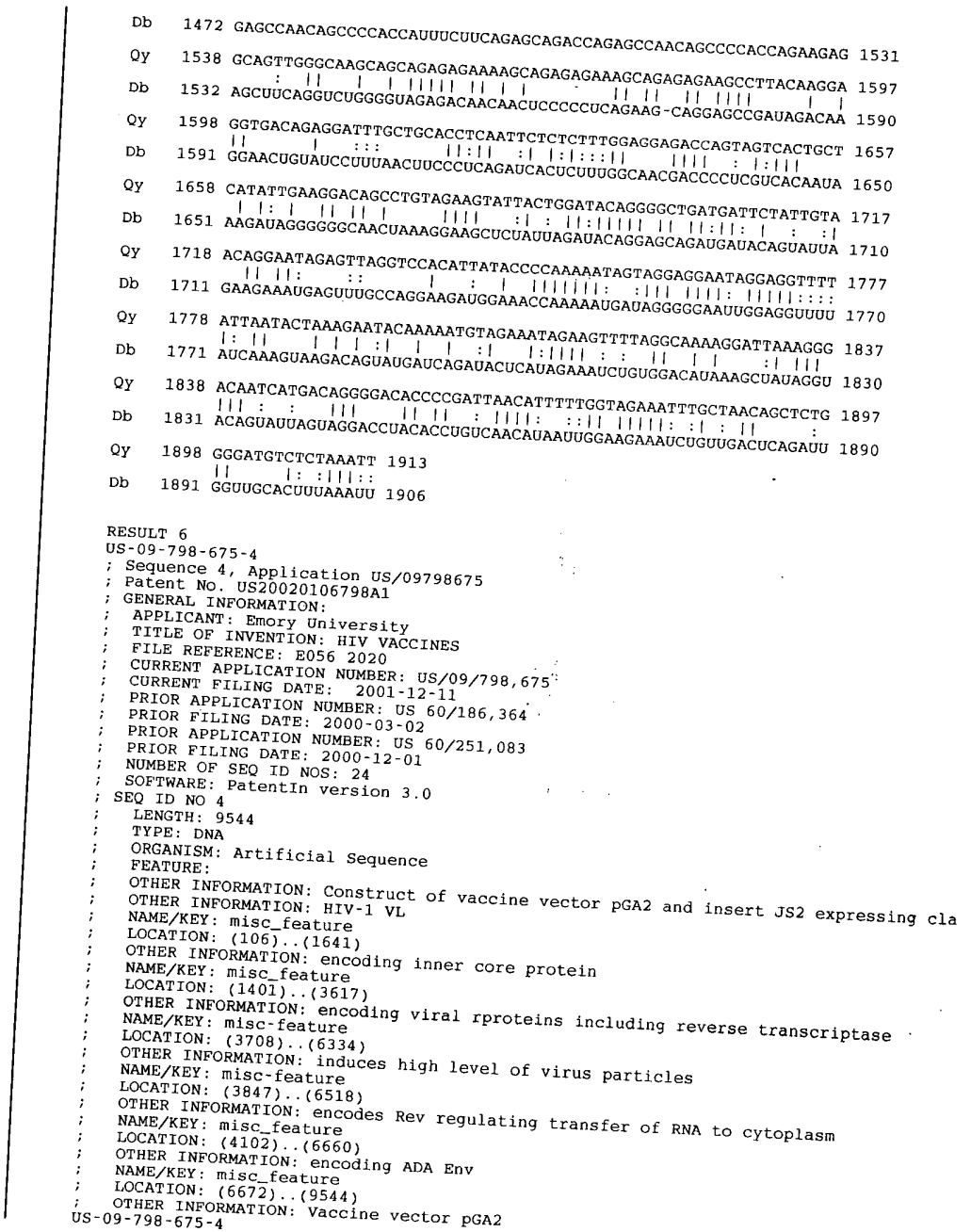




RESULT 5
US-09-943-286-9
; Sequence 9, Application US/09943286
; Patent No. US20020106668A1
; GENERAL INFORMATION:
; APPLICANT: Nunomura, Kiyotada
; TITLE OF INVENTION: POLYNUCLEOTIDE AMPLIFICATION METHOD
; FILE REFERENCE: GP104-02.UT
; CURRENT APPLICATION NUMBER: US/09/943,286
; CURRENT FILING DATE: 2001-08-30
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 8933
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Sequence of the IAC-Bscr pseudo target
; NAME/KEY: mutation
; LOCATION: (4140)...(4159)
; OTHER INFORMATION: Mutated positions: 4140-42, 4145-47, 4152,
; OTHER INFORMATION: 4156-57, 4159
US-09-943-286-9

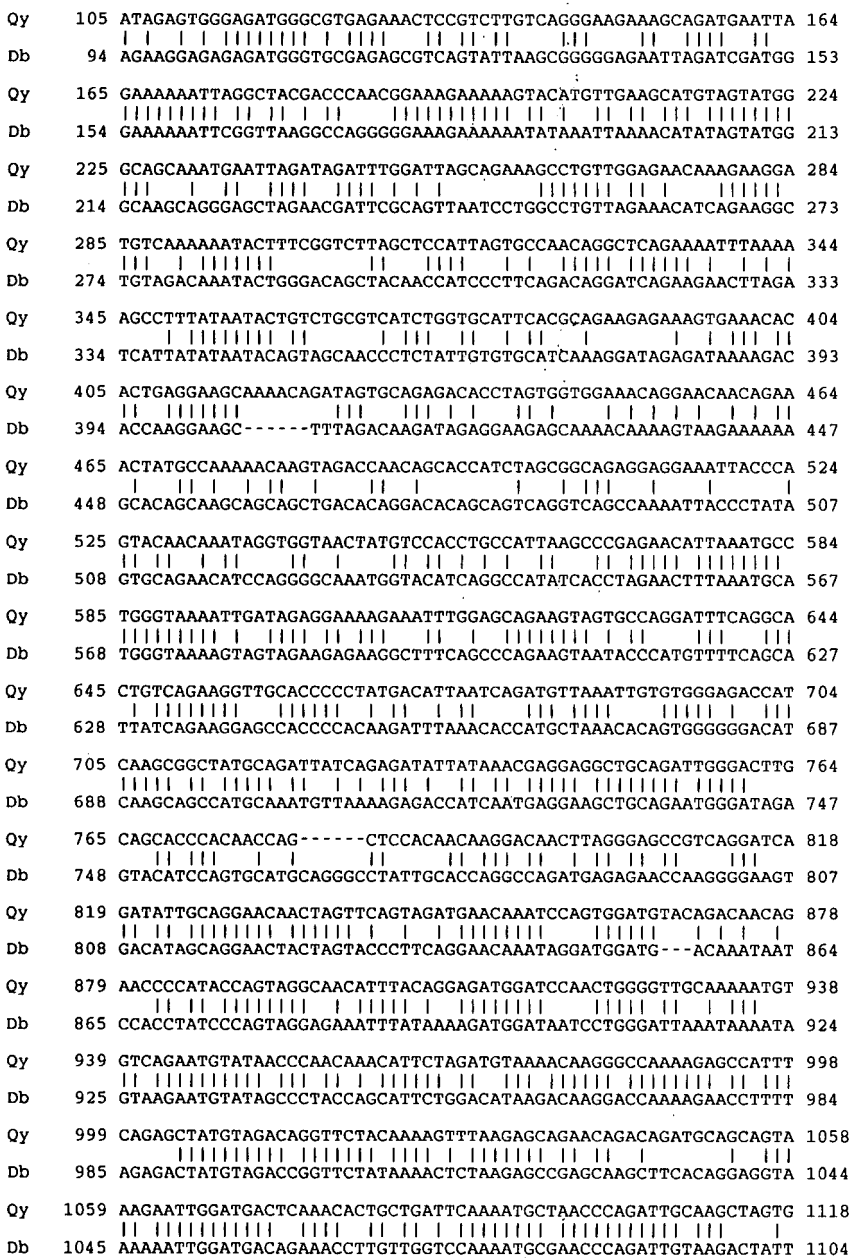
Query Match 26.6%; Score 509.6; DB 10; Length 8933;
Best Local Similarity 45.2%; Pred. No. 9.3e-127;
Matches 821; Conservative 214; Mismatches 764; Indels 17; Gaps 5;

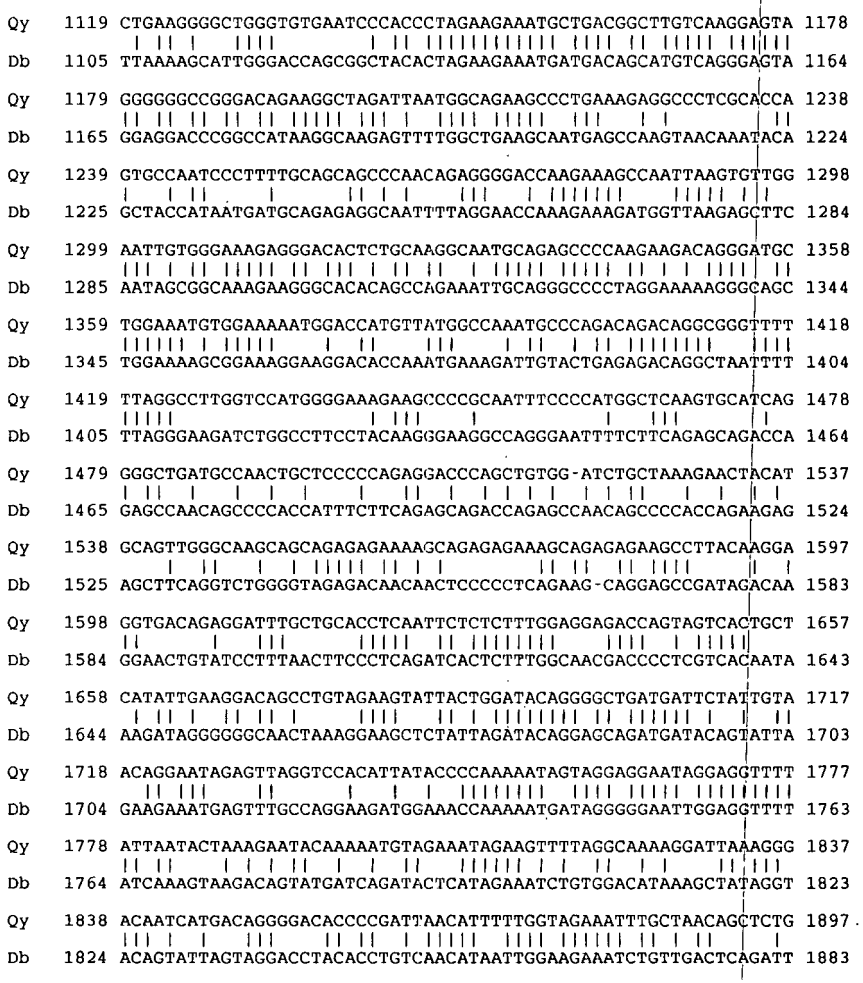




RESULT 6
US-09-798-675-4
; Sequence 4, Application US/09798675
; Patent No. US20020106798A1
; GENERAL INFORMATION:
; APPLICANT: Emory University
; TITLE OF INVENTION: HIV VACCINES
; FILE REFERENCE: E056 2020
; CURRENT APPLICATION NUMBER: US/09/798,675
; CURRENT FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: US 60/186,364
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/251,083
; PRIOR FILING DATE: 2000-12-01
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 9544
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Construct of vaccine vector pGA2 and insert JS2 expressing cla
; OTHER INFORMATION: HIV-1 VL
; NAME/KEY: misc_feature
; LOCATION: (106)..(1641)
; OTHER INFORMATION: encoding inner core protein
; NAME/KEY: misc_feature
; LOCATION: (1401)..(3617)
; OTHER INFORMATION: encoding viral rproteins including reverse transcriptase
; NAME/KEY: misc-feature
; LOCATION: (3708)..(6334)
; OTHER INFORMATION: induces high level of virus particles
; NAME/KEY: misc-feature
; LOCATION: (3847)..(6518)
; OTHER INFORMATION: encodes Rev regulating transfer of RNA to cytoplasm
; NAME/KEY: misc_feature
; LOCATION: (4102)..(6660)
; OTHER INFORMATION: encoding ADA Env
; NAME/KEY: misc_feature
; LOCATION: (6672)..(9544)
; OTHER INFORMATION: Vaccine vector pGA2
US-09-798-675-4

Query Match 26.1%; Score 500; DB 10; Length 9544;
Best Local Similarity 56.7%; Pred. No. 3.7e-124;
Matches 1029; Conservative 0; Mismatches 770; Indels 17; Gaps 5;





RESULT 7
US-09-798-675-5
; Sequence 5, Application US/09798675
; Patent No. US20020106798A1
; GENERAL INFORMATION:
; APPLICANT: Emory University
; TITLE OF INVENTION: HIV VACCINES
; FILE REFERENCE: E056 2020
; CURRENT APPLICATION NUMBER: US/09/798,675
; CURRENT FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: US 60/186,364
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/251,083
; PRIOR FILING DATE: 2000-12-01
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0

```
; SEQ ID NO 5
;   LENGTH: 9918
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Construct of vaccine vector pGA1 and vaccine insert expressing cl
;   OTHER INFORMATION: ade B HIV-1 Gag-Po
;   NAME/KEY: misc_feature
;   LOCATION: (106)..(1641)
;   OTHER INFORMATION: encodes proteins of viral inner core
;   NAME/KEY: misc_feature
;   LOCATION: (1401)..(3617)
;   OTHER INFORMATION: encodes viral proteins but not integrase
;   NAME/KEY: misc_feature
;   LOCATION: (3708)..(5715)
;   OTHER INFORMATION: regulates high-level production of HIV genes
;   NAME/KEY: misc_featre
;   LOCATION: (3847)..(5944)
;   OTHER INFORMATION: encodes protein regulating the transfer of RNA to cytoplasm
;   NAME/KEY: misc_feature
;   LOCATION: (3939)..(4181)
;   OTHER INFORMATION: gene participates in viral assembly and budding
;   NAME/KEY: misc_feature
;   LOCATION: (4099)..(4941)
;   OTHER INFORMATION: encodes truncated form of viral coat protein
;   NAME/KEY: misc_feature
;   LOCATION: (6098)..(9918)
;   OTHER INFORMATION: vaccine vector pGA1
US-09-798-675-5

  Query Match             26.1%;  Score 500;  DB 10;  Length 9918;
  Best Local Similarity   56.7%;  Pred. No. 3.8e-124;
  Matches 1029;  Conservative    0;  Mismatches  770;  Indels   17;  Gaps    5;

Qy    105 ATAGAGTGGGAGATGGGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTA 164
Db     94 AGAAGGAGAGAGATGGGTGCGAGAGCGTCAGTATTAAGCGGGGGAGAATTAGATCGATGG 153

Qy    165 GAAAAAATTAGGCTACGACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGG 224
Db    154 GAAAAAATTCGGTTAAGGCCAGGGGGAAAGAAAAAATATAAATTAAAACATATAGTATGG 213

Qy    225 GCAGCAAATGAATTAGATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGA 284
Db    214 GCAAGCAGGGAGCTAGAACGATTCGCAGTTAATCCTGGCCTGTTAGAAACATCAGAAGGC 273

Qy    285 TGTCAAAAAATACTTTCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAA 344
Db    274 TGTAGACAAATACTGGGACAGCTACAACCATCCCTTCAGACAGGATCAGAAGAACTTAGA 333

Qy    345 AGCCTTTATAATACTGTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACAC 404
Db    334 TCATTATATAATACAGTAGCAACCCTCTATTGTGTGCATCAAAGGATAGAGATAAAAGAC 393

Qy    405 ACTGAGGAAGCAAAACAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAA 464
Db    394 ACCAAGGAAGC------TTTAGACAAGATAGAGGAAGAGCAAAACAAAAGTAAGAAAAAA 447

Qy    465 ACTATGCCAAAAACAAGTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCA 524
Db    448 GCACAGCAAGCAGCAGCTGACACAGGACACAGCAGTCAGGTCAGCCAAAATTACCCTATA 507

Qy    525 GTACAACAAATAGGTGGTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCC 584
Db    508 GTGCAGAACATCCAGGGGCAAATGGTACATCAGGCCATATCACCTAGAACTTTAAATGCA 567

Qy    585 TGGGTAAAATTGATAGAGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCA 644
Db    568 TGGGTAAAAGTAGTAGAAGAGAAGGCTTTCAGCCCAGAAGTAATACCCATGTTTTCAGCA 627

Qy    645 CTGTCAGAAGGTTGCACCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCAT 704
Db    628 TTATCAGAAGGAGCCACCCCACAAGATTTAAACACCATGCTAAACACAGTGGGGGGACAT 687

Qy    705 CAAGCGGCTATGCAGATTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTG 764
Db    688 CAAGCAGCCATGCAAATGTTAAAAGAGACCATCAATGAGGAAGCTGCAGAATGGGATAGA 747

Qy    765 CAGCACCCACAACCAG------CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCA 818
Db    748 GTACATCCAGTGCATGCAGGGCCTATTGCACCAGGCCAGATGAGAGAACCAAGGGGAAGT 807

Qy    819 GATATTGCAGGAACAACTAGTTCAGTAGATGAACAAATCCAGTGGATGTACAGACAACAG 878
Db    808 GACATAGCAGGAACTACTAGTACCCTTCAGGAACAAATAGGATGGATG---ACAAATAAT 864

Qy    879 AACCCCATACCAGTAGGCAACATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGT 938
Db    865 CCACCTATCCCAGTAGGAGAAATTTATAAAAGATGGATAATCCTGGGATTAAATAAAATA 924

Qy    939 GTCAGAATGTATAACCCAACAAACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTT 998
Db    925 GTAAGAATGTATAGCCCTACCAGCATTCTGGACATAAGACAAGGACCAAAAGAACCTTTT 984

Qy    999 CAGAGCTATGTAGACAGGTTCTACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTA 1058
Db    985 AGAGACTATGTAGACCGGTTCTATAAAACTCTAAGAGCCGAGCAAGCTTCACAGGAGGTA 1044

Qy   1059 AAGAATTGGATGACTCAAACACTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTG 1118
Db   1045 AAAAATTGGATGACAGAAACCTTGTTGGTCCAAAATGCGAACCCAGATTGTAAGACTATT 1104

Qy   1119 CTGAAGGGGCTGGGTGTGAATCCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTA 1178
Db   1105 TTAAAAGCATTGGGACCAGCGGCTACACTAGAAGAAATGATGACAGCATGTCAGGGAGTA 1164

Qy   1179 GGGGGGCCGGGACAGAAGGCTAGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCA 1238
Db   1165 GGAGGACCCGGCCATAAGGCAAGAGTTTTGGCTGAAGCAATGAGCCAAGTAACAAATACA 1224

Qy   1239 GTGCCAATCCCTTTTGCAGCAGCCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGG 1298
Db   1225 GCTACCATAATGATGCAGAGAGGCAATTTTAGGAACCAAAGAAAGATGGTTAAGAGCTTC 1284

Qy   1299 AATTGTGGGAAAGAGGGACACTCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGC 1358
Db   1285 AATAGCGGCAAAGAAGGGCACACAGCCAGAAATTGCAGGGCCCCTAGGAAAAAGGGCAGC 1344

Qy   1359 TGGAAATGTGGAAAAATGGACCATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTT 1418
Db   1345 TGGAAAAGCGGAAAGGAAGGACACCAAATGAAAGATTGTACTGAGAGACAGGCTAATTTT 1404

Qy   1419 TTAGGCCTTGGTCCATGGGGAAAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAG 1478
Db   1405 TTAGGGAAGATCTGGCCTTCCTACAAGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCA 1464

Qy   1479 GGGCTGATGCCAACTGCTCCCCCAGAGGACCCAGCTGTGG-ATCTGCTAAAGAACTACAT 1537
Db   1465 GAGCCAACAGCCCCACCATTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAG 1524

Qy   1538 GCAGTTGGGCAAGCAGCAGAGAGAAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGA 1597
Db   1525 AGCTTCAGGTCTGGGGTAGAGACAACAACTCCCCCTCAGAAG-CAGGAGCCGATAGACAA 1583

Qy   1598 GGTGACAGAGGATTTGCTGCACCTCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCT 1657
Db   1584 GGAACTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAACGACCCCTCGTCACAATA 1643

Qy   1658 CATATTGAAGGACAGCCTGTAGAAGTATTACTGGATACAGGGGCTGATGATTCTATTGTA 1717
Db   1644 AAGATAGGGGGGCAACTAAAGGAAGCTCTATTAGATACAGGAGCAGATGATACAGTATTA 1703

Qy   1718 ACAGGAATAGAGTTAGGTCCACATTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTT 1777
Db   1704 GAAGAAATGAGTTTGCCAGGAAGATGGAAACCAAAAATGATAGGGGGAATTGGAGGTTTT 1763
```

```
Qy   1778 ATTAATACTAAAGAATACAAAAATGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGG 1837
          || ||    | | | ||  |  |  ||   |||||| | |  ||  | |    || |||
Db   1764 ATCAAAGTAAGACAGTATGATCAGATACTCATAGAAATCTGTGGACATAAAGCTATAGGT 1823

Qy   1838 ACAATCATGACAGGGGACACCCCGATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTG 1897
          ||| |  |   |||    || ||  | |||||  |||| |||||| || | ||     |
Db   1824 ACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAATCTGTTGACTCAGATT 1883

Qy   1898 GGGATGTCTCTAAATT 1913
          ||     || ||||||
Db   1884 GGTTGCACTTTAAATT 1899


RESULT 8
US-10-003-035-58
; Sequence 58, Application US/10003035
; Patent No. US20020155127A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Danher
;  TITLE OF INVENTION: GENETIC VACCINE AGAINST HUMAN IMMUNODEFICIENCY VIRUS
;  FILE REFERENCE: 22488-712
;  CURRENT APPLICATION NUMBER: US/10/003,035
;  CURRENT FILING DATE:  2001-11-01
;  PRIOR APPLICATION NUMBER: 09/585,599
;  PRIOR FILING DATE: 2000-06-02
;  PRIOR APPLICATION NUMBER: PCT/US01/18238
;  PRIOR FILING DATE: 2001-06-04
;  NUMBER OF SEQ ID NOS: 75
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 58
;   LENGTH: 1800
;   TYPE: DNA
;   ORGANISM: Artificial sequence
;   FEATURE:
;   OTHER INFORMATION: Gag-PI
US-10-003-035-58

  Query Match             26.0%;  Score 498.2;  DB 9;  Length 1800;
  Best Local Similarity   56.7%;  Pred. No. 4.4e-124;
  Matches 1024;  Conservative    0;  Mismatches  763;  Indels   18;  Gaps    5;

Qy    117 ATGGGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGG 176
          |||||  | ||||   || || ||    |||    ||  ||||  ||  |||||||||| ||
Db      1 ATGGGTGCGAGAGCGTCAGTATTAAGCGGGGGAGAATTAGATCGATGGGAAAAAATTCGG 60

Qy    177 CTACGACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAA 236
          || | ||    ||||||||||| || |  || || ||| ||||||||||    |  ||
Db     61 TTAAGGCCAGGGGGAAAGAAAAAATATAAATTAAAACATATAGTATGGGCAAGCAGGGAG 120

Qy    237 TTAGATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATA 296
          ||||  |||| | | |        |||||| || |    |||||| ||| | |||||
Db    121 CTAGAACGATTCGCAGTTAATCCTGGCCTGTTAGAAACATCAGAAGGCTGTAGACAAATA 180

Qy    297 CTTTCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAAT 356
          ||        ||   ||||   | |  ||||| |||||| |  | | |    | ||||||
Db    181 CTGGGACAGCTACAACCATCCCTTCAGACAGGATCAGAAGAACTTAGATCATTATATAAT 240

Qy    357 ACTGTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCA 416
          || ||      | |||  ||  | ||   |        |  | ||| ||||  |||||||
Db    241 ACAGTAGCAACCCTCTATTGTGTGCATCAAAGGATAGAGATAAAAGACACCAAGGAAGC- 299

Qy    417 AAACAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAA 476
              |||   ||| | |   || |    | | | | |  | | || |   || | |
Db    300 -----TTTAGACAAGATAGAGGAAGAGCAAAACAAAAGTAAGAAAAAAGCACAGCAAGCA 354

Qy    477 ACAAGTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATA 536
          ||  |    ||  |        |   | |||   |     |       ||| ||  | ||
Db    355 GCAGCTGACACAGGACACAGCAGTCAGGTCAGCCAAAATTACCCTATAGTGCAGAACATC 414

Qy    537 GGTGGTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTG 596
          ||  |    || || | |  ||   || ||||| |||||||| |||||||| |
Db    415 CAGGGGCAAATGGTACATCAGGCCATATCACCTAGAACTTTAAATGCATGGGTAAAAGTA 474

Qy    597 ATAGAGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGT 656
          |||| || |||   ||  |  |||||||| | ||    |||   ||| | ||||||||
Db    475 GTAGAAGAGAAGGCTTTCAGCCCAGAAGTAATACCCATGTTTTCAGCATTATCAGAAGGA 534

Qy    657 TGCACCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATG 716
          ||||||  | ||  | ||   ||| ||||    ||||| |  |||||||| || |||
Db    535 GCCACCCCACAAGATTTAAACACCATGCTAAACACAGTGGGGGGACATCAAGCAGCCATG 594

Qy    717 CAGATTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAA 776
          || ||  | | |||  |  || || ||||| |||||||| |||||       || |||
Db    595 CAAATGTTAAAAGAGACCATCAATGAGGAAGCTGCAGAATGGGATAGAGTACATCCAGTG 654

Qy    777 CCAG------CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGA 830
          |  |      ||     || ||| || | || || ||    |||   || || |||||||
Db    655 CATGCAGGGCCTATTGCACCAGGCCAGATGAGAGAACCAAGGGGAAGTGACATAGCAGGA 714

Qy    831 ACAACTAGTTCAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCA 890
          || |||||| |  |  | ||||||||    ||||||   | | |  |    || || |||
Db    715 ACTACTAGTACCCTTCAGGAACAAATAGGATGGATG---ACAAATAATCCACCTATCCCA 771

Qy    891 GTAGGCAACATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTAT 950
          |||||  | |||||  |  ||||||||    |||||  ||  | |||   || |||||||||
Db    772 GTAGGAGAAATTTATAAAAGATGGATAATCCTGGGATTAAATAAAATAGTAAGAATGTAT 831

Qy    951 AACCCAACAAACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTA 1010
          | ||| || | |||||| ||  ||| |||||| |||||||| || |||     ||||||
Db    832 AGCCCTACCAGCATTCTGGACATAAGACAAGGACCAAAAGAACCTTTTAGAGACTATGTA 891

Qy   1011 GACAGGTTCTACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATG 1070
          ||| ||||||| |||| | ||||||| || ||  |      |  ||||| |||||||||
Db    892 GACCGGTTCTATAAAACTCTAAGAGCCGAGCAAGCTTCACAGGAGGTAAAAAATTGGATG 951

Qy   1071 ACTCAAACACTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTG 1130
          ||  ||||  || || | |||||||| |||||||||| |||    |  | || |  ||
Db    952 ACAGAAACCTTGTTGGTCCAAAATGCGAACCCAGATTGTAAGACTATTTTAAAAGCATTG 1011

Qy   1131 GGTGTGAATCCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGA 1190
          ||        | || |||||||||||| |||| || ||||| |||||||| || || ||
Db   1012 GGACCAGCGGCTACACTAGAAGAAATGATGACAGCATGTCAGGGAGTAGGAGGACCCGGC 1071

Qy   1191 CAGAAGGCTAGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCT 1250
          || ||||| ||| |  |||| |||||  |||   | |        |||   | ||
Db   1072 CATAAGGCAAGAGTTTTGGCTGAAGCAATGAGCCAAGTAACAAATACAGCTACCATAATG 1131

Qy   1251 TTTGCAGCAGCCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAA 1310
          |       || | |    |||   | |||||||    |||||||||  |||||||| |||
Db   1132 ATGCAGAGAGGCAATTTTAGGAACCAAAGAAAGATGGTTAAGTGTTTCAATTGTGGCAAA 1191

Qy   1311 GAGGGACACTCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGA 1370
          || || ||| | || ||   | ||||| ||||| || | |  |||| || ||||||||||
Db   1192 GAAGGGCACACAGCCAGAAATTGCAGGGCCCCTAGGAAAAAGGGCTGTTGGAAATGTGGA 1251

Qy   1371 AAAATGGACCATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGT 1430
          ||    | ||    |||    | ||  | || |||||||   ||| ||  |    || |
Db   1252 AAGGAAGGACACCAAATGAAAGATTGTACTGAGAGACAGGCTAATTTCTTTAGGGAAGAT 1311

Qy   1431 C--CATGGGGAAAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGC 1488
          |     |  | |    || |       |      ||   | ||| | |   | ||
Db   1312 CTGGCCTTCCTACAAGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGC 1371

Qy   1489 CAACTGCTCCCCCAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCA 1548
          |     |      |  | ||   |  | |  |  |  |  |    |   ||
Db   1372 CCCACCATTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTCAGGTC 1431

Qy   1549 AGCAGCAGAGAGAAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGG 1608
          |  | |||||  || | |         || ||  || ||||     |  |||       |
Db   1432 TGGGGTAGAGACAACAACTCCCCCTCAGAAGC-AGGAGCCGATAGACAAGGAACTGTATC 1490

Qy   1609 ATTTGCTGCACCTCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGG 1668
          |||      |||||  || ||||||||    ||||  | |||||     | || ||  ||
Db   1491 CTTTAACTTCCCTCAGATCACTCTTTGGCAACGACCCCTCGTCACAATAAAGATAGGGGG 1550

Qy   1669 ACAGCCTGTAGAAGTATTACTGGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGA 1728
```

```
Db   1551 GCAACTAAAGGAAGCTCTATTAGATACAGGAGCAGATGATACAGTATTAGAAGAAATGAG 1610

Qy   1729 GTTAGGTCCACATTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAA 1788
          ||       |   |  | ||||||||  |||| |||||  |||||||||||  ||    |
Db   1611 TTTGCCAGGAAGATGGAAACCAAAAATGATAGGGGGAATTGGAGGTTTTATCAAAGTAAG 1670

Qy   1789 AGAATACAAAAATGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGAC 1848
          | | ||  |  |  ||   |||||| | |  ||  | |    || ||| ||| | |
Db   1671 ACAGTATGATCAGATACTCATAGAAATCTGTGGACATAAAGCTATAGGTACAGTATTAGT 1730

Qy   1849 AGGGGACACCCCGATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCT 1908
          |||    || ||  | |||||  |||| |||||| || | ||     | ||     || |
Db   1731 AGGACCTACACCTGTCAACATAATTGGAAGAAATCTGTTGACTCAGATTGGTTGCACTTT 1790

Qy   1909 AAATT 1913
          |||||
Db   1791 AAATT 1795


RESULT 9
US-09-737-190A-1
; Sequence 1, Application US/09737190A
; Patent No. US20020102545A1
; GENERAL INFORMATION:
;  APPLICANT: Shibuya, Tetsuo
;  TITLE OF INVENTION: A Method for Changing a Target Array, a Method for Analyzing
;  TITLE OF INVENTION:  a Structure, and an Apparatus, a Storage Medium and a
;  TITLE OF INVENTION:  Transmission Medium Therefor
;  FILE REFERENCE: JP919990270US1 (14043)
;  CURRENT APPLICATION NUMBER: US/09/737,190A
;  CURRENT FILING DATE:  2000-12-14
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 9719
;   TYPE: DNA
;   ORGANISM: Human iImmunodeficiency virus type 1
US-09-737-190A-1

  Query Match             25.4%;  Score 485.4;  DB 10;  Length 9719;
  Best Local Similarity   56.6%;  Pred. No. 3.1e-120;
  Matches 1028;  Conservative    0;  Mismatches  736;  Indels   51;  Gaps    5;

Qy     105 ATAGAGTGGGAGATGGGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTA 164
           | |  | | |||||||| | ||||   || || ||   |||   ||  ||||  ||
Db     778 AGAAGGAGAGAGATGGGTGCGAGAGCGTCAGTATTAAGCGGGGGAGAATTAGATCGATGG 837

Qy     165 GAAAAAATTAGGCTACGACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGG 224
           ||||||||| || || | ||    |||||||||| || |  || || ||| ||||||||
Db     838 GAAAAAATTCGGTTAAGGCCAGGGGGAAAGAAAAAATATAAATTAAAACATATAGTATGG 897

Qy     225 GCAGCAAATGAATTAGATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGA 284
           |||   |  ||  ||||  |||| | | |       ||||||| || |    ||||||
Db     898 GCAAGCAGGGAGCTAGAACGATTCGCAGTTAATCCTGGCCTGTTAGAAACATCAGAAGGC 957

Qy     285 TGTCAAAAAATACTTTCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAA 344
           |||  | |||||||        ||   ||||   | |  ||||| |||||| |  | | |
Db     958 TGTAGACAAATACTGGGACAGCTACAACCATCCCTTCAGACAGGATCAGAAGAACTTAGA 1017

Qy     345 AGCCTTTATAATACTGTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACAC 404
           | |||||||| ||      | |||  ||  | ||   |       |  | ||| ||
Db    1018 TCATTATATAATACAGTAGCAACCCTCTATTGTGTGCATCAAAGGATAGAGATAAAAGAC 1077

Qy     405 ACTGAGGAAGCAAAACAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAA 464
           ||  |||||||          |||   ||| | |   || |    | | | |  | | ||
Db    1078 ACCAAGGAAGC------TTTAGACAAGATAGAGGAAGAGCAAAACAAAAGTAAGAAAAAA 1131

Qy     465 ACTATGCCAAAAACAAGTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCA 524
           |  || | | || |    ||  |         |  | |||   |    |       |
Db    1132 GCACAGCAAGCAGCAGCTGACACAGGACACAGCAATCAGGTCAGCCAAAATTACCCTATA 1191

Qy     525 GTACAACAAATAGGTGGTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCC 584
```

```
Db    1192 GTGCAGAACATCCAGGGGCAAATGGTACATCAGGCCATATCACCTAGAACTTTAAATGCA 1251

Qy     585 TGGGTAAAATTGATAGAGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCA 644
           |||||||||  |  |||| || |||    ||  |  |||||||  | ||    |||   |||
Db    1252 TGGGTAAAAGTAGTAGAAGAGAAGGCTTTCAGCCCAGAAGTGATACCCATGTTTTCAGCA 1311

Qy     645 CTGTCAGAAGGTTGCACCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCAT 704
           | |||||||||   ||||||  | ||  | ||    ||| ||||    |||||  |  |||
Db    1312 TTATCAGAAGGAGCCACCCCACAAGATTTAAACACCATGCTAAACACAGTGGGGGGACAT 1371

Qy     705 CAAGCGGCTATGCAGATTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTG 764
           |||||  || |||||  ||   |  | |||  |   ||  ||  |||||  |||||||| |||||
Db    1372 CAAGCAGCCATGCAAATGTTAAAAGAGACCATCAATGAGGAAGCTGCAGAATGGGATAGA 1431

Qy     765 CAGCACCCACAACCAG------CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCA 818
             |||  |||    |  |       ||      ||  |||  ||   |  ||  ||  ||    |||
Db    1432 GTGCATCCAGTGCATGCAGGGCCTATTGCACCAGGCCAGATGAGAGAACCAAGGGGAAGT 1491

Qy     819 GATATTGCAGGAACAACTAGTTCAGTAGATGAACAAATCCAGTGGATGTACAGACAACAG 878
           ||  ||  ||||||||  ||||||  |   |   |  ||||||||    ||||||    | | |  |
Db    1492 GACATAGCAGGAACTACTAGTACCCTTCAGGAACAAATAGGATGGATG---ACAAATAAT 1548

Qy     879 AACCCCATACCAGTAGGCAACATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGT 938
             ||  ||  |||||||||   |  |||||  |   ||||||||     |||||  ||   | |||
Db    1549 CCACCTATCCCAGTAGGAGAAATTTATAAAAGATGGATAATCCTGGGATTAAATAAAATA 1608

Qy     939 GTCAGAATGTATAACCCAACAAACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTT 998
           ||  ||||||||||  |||  ||  |  ||||||  ||   |||  ||||||  |||||  ||  ||| |||
Db    1609 GTAAGAATGTATAGCCCTACCAGCATTCTGGACATAAGACAAGGACCAAAGGAACCCTTT 1668

Qy     999 CAGAGCTATGTAGACAGGTTCTACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTA 1058
             |||||||||  |||||||  ||||  |  |||||||  ||  ||   |       |  |||
Db    1669 AGAGACTATGTAGACCGGTTCTATAAAACTCTAAGAGCCGAGCAAGCTTCACAGGAGGTA 1728

Qy    1059 AAGAATTGGATGACTCAAACACTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTG 1118
           ||  ||||||||||||  ||||  ||  ||  |  |||||||  ||||||||||  |||     |
Db    1729 AAAAATTGGATGACAGAAACCTTGTTGGTCCAAAATGCGAACCCAGATTGTAAGACTATT 1788

Qy    1119 CTGAAGGGGCTGGGTGTGAATCCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTA 1178
           | ||  |   ||||          | ||  ||||||||||| |||| ||  ||||| ||||||
Db    1789 TTAAAAGCATTGGGACCAGCGGCTACACTAGAAGAAATGATGACAGCATGTCAGGGAGTA 1848

Qy    1179 GGGGGGCCGGGACAGAAGGCTAGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCA 1238
           ||  ||  ||  ||  ||  |||||  |||  |   ||||  |||||   |||    | |          ||
Db    1849 GGAGGACCCGGCCATAAGGCAAGAGTTTTGGCTGAAGCAATGAGCCAAGTAACAAATTCA 1908

Qy    1239 GTGCCAATCCCTTTTGCAGCAGCCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGG 1298
           |   | ||         |      ||  |  |    |||   | |||||||    |||||||||
Db    1909 GCTACCATAATGATGCAGAGAGGCAATTTTAGGAACCAAAGAAAGATTGTTAAGTGTTTC 1968

Qy    1299 AATTGTGGGAAAGAGGGACACTCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGC 1358
           ||||||||  |||||  ||  |||  |  ||  ||   |  |||||  |||||  ||  | |  ||
Db    1969 AATTGTGGCAAAGAAGGGCACACAGCCAGAAATTGCAGGGCCCCTAGGAAAAAGGGCTGT 2028

Qy    1359 TGGAAATGTGGAAAAATGGACCATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTT 1418
           ||||||||||||||     |  ||     |||     | ||  |  || ||||||||    ||||
Db    2029 TGGAAATGTGGAAAGGAAGGACACCAAATGAAAGATTGTACTGAGAGACAGGCTAATTTT 2088

Qy    1419 TTAGGCCTTGGTCCATGGGGAAAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAG 1478
           |||||                | |||    |      ||   |  |    | ||
Db    2089 TTAGGGAAGATCTGGCCTTCCTACAAGGGAAG------GCCAGGGAATTTTCTTCAGAGC 2142

Qy    1479 GGGCTGATGCCAACTGCTCCCCCAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATG 1538
           | |    ||||||  || || |||||  ||          |  |  | || |||     |
Db    2143 AGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTCAGGTCTGGGGTAGAGACAACAACT 2202

Qy    1539 CAGTTGGGCAAGCAGCAGAGAGAAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAG 1598
           |       ||||                                || ||||     |  ||
Db    2203 CCCCCTCAGAAGC------------------------------AGGAGCCGATAGACAAG 2232

Qy    1599 GTGACAGAGGATTTGCTGCACCTCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTC 1658
           |       |   |||      |||||  || ||||||||    ||||  | |||||
```

```
Db   2233 GAACTGTATCCTTTAACTTCCCTCAGGTCACTCTTTGGCAACGACCCCTCGTCACAATAA 2292

Qy   1659 ATATTGAAGGACAGCCTGTAGAAGTATTACTGGATACAGGGGCTGATGATTCTATTGTAA 1718

Db   2293 AGATAGGGGGGCAACTAAAGGAAGCTCTATTAGATACAGGAGCAGATGATACAGTATTAG 2352

Qy   1719 CAGGAATAGAGTTAGGTCCACATTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTA 1778

Db   2353 AAGAAATGAGTTTGCCAGGAAGATGGAAACCAAAAATGATAGGGGGAATTGGAGGTTTTA 2412

Qy   1779 TTAATACTAAAGAATACAAAAATGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGA 1838

Db   2413 TCAAAGTAAGACAGTATGATCAGATACTCATAGAAATCTGTGGACATAAAGCTATAGGTA 2472

Qy   1839 CAATCATGACAGGGGACACCCCGATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTGG 1898

Db   2473 CAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAATCTGTTGACTCAGATTG 2532

Qy   1899 GGATGTCTCTAAATT 1913

Db   2533 GTTGCACTTTAAATT 2547


RESULT 10
US-09-999-183-1
; Sequence 1, Application US/09999183
; Patent No. US20020147169A1
; GENERAL INFORMATION:
;  APPLICANT: MITROPHANOUS, et al
;  TITLE OF INVENTION: In Vivo Selection Method
;  FILE REFERENCE: 674523-2009
;  CURRENT APPLICATION NUMBER: US/09/999,183
;  CURRENT FILING DATE:  2001-11-29
;  PRIOR APPLICATION NUMBER: PCT/GB00/02136
;  PRIOR FILING DATE: 2000-06-02
;  PRIOR APPLICATION NUMBER: 9912965.2
;  PRIOR FILING DATE: 1999-06-03
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: SeqWin99
; SEQ ID NO 1
;   LENGTH: 4307
;   TYPE: DNA
;   ORGANISM: Human immunodeficiency virus type I
US-09-999-183-1

  Query Match             25.2%;  Score 481.4;  DB 10;  Length 4307;
  Best Local Similarity   56.6%;  Pred. No. 2.4e-119;
  Matches 1021;  Conservative    0;  Mismatches  731;  Indels   51;  Gaps    5;

Qy    117 ATGGGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGG 176

Db      1 ATGGGTGCGAGAGCGTCAGTATTAAGCGGGGGAGAATTAGATCGATGGGAAAAAATTCGG 60

Qy    177 CTACGACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAA 236

Db     61 TTAAGGCCAGGGGGAAAGAAAAAATATAAATTAAAACATATAGTATGGGCAAGCAGGGAG 120

Qy    237 TTAGATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATA 296

Db    121 CTAGAACGATTCGCAGTTAATCCTGGCCTGTTAGAAACATCAGAAGGCTGTAGACAAATA 180

Qy    297 CTTTCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAAT 356

Db    181 CTGGGACAGCTACAACCATCCCTTCAGACAGGATCAGAAGAACTTAGATCATTATATAAT 240

Qy    357 ACTGTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCA 416

Db    241 ACAGTAGCAACCCTCTATTGTGTGCATCAAAGGATAGAGATAAAAGACACCAAGGAAGC- 299

Qy    417 AAACAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAA 476

Db    300 -----TTTAGACAAGATAGAGGAAGAGCAAAACAAAAGTAAGAAAAAAGCACAGCAAGCA 354

Qy    477 ACAAGTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATA 536

Db    355 GCAGCTGACACAGGACACAGCAATCAGGTCAGCCAAAATTACCCTATAGTGCAGAACATC 414

Qy    537 GGTGGTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTG 596

Db    415 CAGGGGCAAATGGTACATCAGGCCATATCACCTAGAACTTTAAATGCATGGGTAAAAGTA 474

Qy    597 ATAGAGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGT 656

Db    475 GTAGAAGAGAAGGCTTTCAGCCCAGAAGTGATACCCATGTTTTCAGCATTATCAGAAGGA 534

Qy    657 TGCACCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATG 716

Db    535 GCCACCCCACAAGATTTAAACACCATGCTAAACACAGTGGGGGGACATCAAGCAGCCATG 594

Qy    717 CAGATTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAA 776

Db    595 CAAATGTTAAAAGAGACCATCAATGAGGAAGCTGCAGAATGGGATAGAGTGCATCCAGTG 654

Qy    777 CCAG------CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGA 830

Db    655 CATGCAGGGCCTATTGCACCAGGCCAGATGAGAGAACCAAGGGGAAGTGACATAGCAGGA 714

Qy    831 ACAACTAGTTCAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCA 890

Db    715 ACTACTAGTACCCTTCAGGAACAAATAGGATGGATG---ACAAATAATCCACCTATCCCA 771

Qy    891 GTAGGCAACATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTAT 950

Db    772 GTAGGAGAAATTTATAAAAGATGGATAATCCTGGGATTAAATAAAATAGTAAGAATGTAT 831

Qy    951 AACCCAACAAACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTA 1010

Db    832 AGCCCTACCAGCATTCTGGACATAAGACAAGGACCAAAGGAACCCTTTAGAGACTATGTA 891

Qy   1011 GACAGGTTCTACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATG 1070

Db    892 GACCGGTTCTATAAAACTCTAAGAGCCGAGCAAGCTTCACAGGAGGTAAAAAATTGGATG 951

Qy   1071 ACTCAAACACTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTG 1130

Db    952 ACAGAAACCTTGTTGGTCCAAAATGCGAACCCAGATTGTAAGACTATTTTAAAAGCATTG 1011

Qy   1131 GGTGTGAATCCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGA 1190

Db   1012 GGACCAGCGGCTACACTAGAAGAAATGATGACAGCATGTCAGGGAGTAGGAGGACCCGGC 1071

Qy   1191 CAGAAGGCTAGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCT 1250

Db   1072 CATAAGGCAAGAGTTTTGGCTGAAGCAATGAGCCAAGTAACAAATTCAGCTACCATAATG 1131

Qy   1251 TTTGCAGCAGCCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAA 1310

Db   1132 ATGCAGAGAGGCAATTTTAGGAACCAAAGAAAGATTGTTAAGTGTTTCAATTGTGGCAAA 1191

Qy   1311 GAGGGACACTCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGA 1370

Db   1192 GAAGGGCACACAGCCAGAAATTGCAGGGCCCCTAGGAAAAGGGCTGTTGGAAATGTGGA 1251

Qy   1371 AAAATGGACCATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGGCCTTGGT 1430

Db   1252 AAGGAAGGACACCAAATGAAAGATTGTACTGAGAGACAGGCTAATTTTTTAGGGAAGATC 1311

Qy   1431 CCATGGGGAAAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGGGGCTGATGCCA 1490

Db   1312 TGGCCTTCCTACAAGGGAAG------GCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCA 1365

Qy   1491 ACTGCTCCCCCAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAG 1550

Db   1366 ACAGCCCCACCAGAAGAGAGCTTCAGGTCTGGGGTAGAGACAACAACTCCCCCTCAGAAG 1425

Qy   1551 CAGCAGAGAGAAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGGTGACAGAGGAT 1610
```

```
Db   1426 C----------------------------AGGAGCCGATAGACAAGGAACTGTATCCT 1455

Qy   1611 TTGCTGCACCTCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGAC 1670
Db   1456 TTAACTTCCCTCAGGTCACTCTTTGGCAACGACCCCTCGTCACAATAAAGATAGGGGGGC 1515

Qy   1671 AGCCTGTAGAAGTATTACTGGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGT 1730
Db   1516 AACTAAAGGAAGCTCTATTAGATACAGGAGCAGATGATACAGTATTAGAAGAAATGAGTT 1575

Qy   1731 TAGGTCCACATTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAG 1790
Db   1576 TGCCAGGAAGATGGAAACCAAAAATGATAGGGGGAATTGGAGGTTTTATCAAAGTAAGAC 1635

Qy   1791 AATACAAAAATGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAG 1850
Db   1636 AGTATGATCAGATACTCATAGAAATCTGTGGACATAAAGCTATAGGTACAGTATTAGTAG 1695

Qy   1851 GGGACACCCCGATTAACATTTTTGGTAGAAATTTGCTAACAGCTCTGGGGATGTCTCTAA 1910
Db   1696 GACCTACACCTGTCAACATAATTGGAAGAAATCTGTTGACTCAGATTGGTTGCACTTTAA 1755

Qy   1911 ATT 1913
Db   1756 ATT 1758


RESULT 11
US-09-872-733-6
; Sequence 6, Application US/09872733
; Patent No. US20010036655A1
; GENERAL INFORMATION:
;  APPLICANT: The Government of the United States of America, as
;  TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, SIV GAG AND
;  TITLE OF INVENTION:  SIV ENV GENES
;  FILE REFERENCE: 2026-4287US1 HIV GAG/POL,SIV GAG & ENV
;  CURRENT APPLICATION NUMBER: US/09/872,733
;  CURRENT FILING DATE:  2001-06-01
;  PRIOR APPLICATION NUMBER: PCT/US00/34985
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/173,036
;  PRIOR FILING DATE: 1999-12-23
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 8366
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  DNA sequence
;   OTHER INFORMATION: of the construct pCMVgagpolBNKan containing a CMV
;   OTHER INFORMATION: promoter, a HIV gag/pol gene and a kanamycin
;   OTHER INFORMATION: resistance gene
US-09-872-733-6

  Query Match           23.8%;  Score 455;  DB 10;  Length 8366;
  Best Local Similarity  55.8%;  Pred. No. 4.3e-112;
  Matches 1000;  Conservative    0;  Mismatches  740;  Indels   51;  Gaps    5;

Qy    106 TAGAGTGGGAGATGGGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAG 165
Db    759 TCGACAGAGAGATGGGTGCGAGAGCGTCAGTATTAAGCGGGGGAGAATTAGATCGATGGG 818

Qy    166 AAAAAATTAGGCTACGACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGG 225
Db    819 AAAAAATTCGGTTAAGGCCAGGGGGAAAGAAGAAGTACAAGCTAAAGCACATCGTATGGG 878

Qy    226 CAGCAAATGAATTAGATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGAT 285
Db    879 CAAGCAGGGAGCTAGAACGATTCGCAGTTAATCCTGGCCTGTTAGAAACATCAGAAGGCT 938

Qy    286 GTCAAAAAATACTTTCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAA 345

Db    939 GTAGACAAATACTGGGACAGCTACAACCATCCCTTCAGACAGGATCAGAGGAGCTTCGAT 998

Qy    346 GCCTTTATAATACTGTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACA 405
Db    999 CACTATACAACACAGTAGCAACCCTCTATTGTGTGCACCAGCGGATCGAGATCAAGGACA 1058

Qy    406 CTGAGGAAGCAAAACAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAA 465
Db   1059 CCAAGGAAGC------TTTAGACAAGATAGAGGAAGAGCAAAACAAGTCCAAGAAGAAGG 1112

Qy    466 CTATGCCAAAAACAAGTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAG 525
Db   1113 CCCAGCAGGCAGCAGCTGACACAGGACACAGCAATCAGGTCAGCCAAAATTACCCTATAG 1172

Qy    526 TACAACAAATAGGTGGTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCT 585
Db   1173 TGCAGAACATCCAGGGGCAAATGGTACATCAGGCCATATCACCTAGAACTTTAAATGCAT 1232

Qy    586 GGGTAAAATTGATAGAGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCAC 645
Db   1233 GGGTAAAAGTAGTAGAAGAGAAGGCTTTCAGCCCAGAAGTGATACCCATGTTTTCAGCAT 1292

Qy    646 TGTCAGAAGGTTGCACCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATC 705
Db   1293 TATCAGAAGGAGCCACCCCACAGGACCTGAACACGATGTTGAACACCGTGGGGGGACATC 1352

Qy    706 AAGCGGCTATGCAGATTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGC 765
Db   1353 AAGCAGCCATGCAAATGTTAAAAGAGACCATCAATGAGGAAGCTGCAGAATGGGATAGAG 1412

Qy    766 AGCACCCACAACCAG------CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATCAG 819
Db   1413 TGCATCCAGTGCATGCAGGGCCTATTGCACCAGGCCAGATGAGAGAACCAAGGGGAAGTG 1472

Qy    820 ATATTGCAGGAACAACTAGTTCAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGA 879
Db   1473 ACATAGCAGGAACTACTAGTACCCTTCAGGAACAAATAGGATGGATG---ACAAATAATC 1529

Qy    880 ACCCCATACCAGTAGGCAACATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTG 939
Db   1530 CACCTATCCCAGTAGGAGAGATCTACAAGAGGTGGATAATCCTGGGATTGAACAAGATCG 1589

Qy    940 TCAGAATGTATAACCCAACAAACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTC 999
Db   1590 TGAGGATGTATAGCCCTACCAGCATTCTGGACATAAGACAAGGACCAAAGGAACCCTTTA 1649

Qy   1000 AGAGCTATGTAGACAGGTTCTACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAA 1059
Db   1650 GAGACTATGTAGACCGGTTCTATAAAACTCTAAGAGCTGAGCAAGCTTCACAGGAGGTAA 1709

Qy   1060 AGAATTGGATGACTCAAACACTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGC 1119
Db   1710 AAAATTGGATGACAGAAACCTTGTTGGTCCAAAATGCGAACCCAGATTGTAAGACCATCC 1769

Qy   1120 TGAAGGGGCTGGGTGTGAATCCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAG 1179
Db   1770 TGAAGGCTCTCGGCCCAGCGGCTACACTAGAAGAAATGATGACAGCATGTCAGGGAGTAG 1829

Qy   1180 GGGGGCCGGGACAGAAGGCTAGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAG 1239
Db   1830 GAGGACCCGGCCATAAGGCAAGAGTTTTGGCCGAGGCGATGAGCCAGGTGACGAACTCGG 1889

Qy   1240 TGCCAATCCCTTTTGCAGCAGCCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGA 1299
Db   1890 CGACCATAATGATGCAGAGAGGCAACTTCCGGAACCAGCGGAAGATCGTCAAGTGCTTCA 1949

Qy   1300 ATTGTGGGAAAGAGGGACACTCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCT 1359
Db   1950 ATTGTGGCAAAGAAGGGCACACCGCCAGGAACTGCCGGGCCCCCCGGAAGAAGGGCTGTT 2009

Qy   1360 GGAAATGTGGAAAAATGGACCATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTT 1419
```

```
Db   2010 GGAAATGTGGAAAGGAAGGACACCAAATGAAAGATTGTACTGAGAGACAGGCTAATTTTT 2069

Qy   1420 TAGGCCTTGGTCCATGGGGAAAGAAGCCCCGCAATTTCCCCATGGCTCAAGTGCATCAGG 1479
Db   2070 TAGGGAAGATCTGGCCTTCCTACAAGGGAAG------GCCAGGGAATTTTCTTCAGAGCA 2123

Qy   1480 GGCTGATGCCAACTGCTCCCCCAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGC 1539
Db   2124 GACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTCAGGTCTGGGGTAGAGACAACAACTC 2183

Qy   1540 AGTTGGGCAAGCAGCAGAGAGAAAAGCAGAGAGAAAGCAGAGAGAAGCCTTACAAGGAGG 1599
Db   2184 CCCCTCAGAAGC----------------------------AGGAGCCGATAGACAAGG 2213

Qy   1600 TGACAGAGGATTTGCTGCACCTCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCA 1659
Db   2214 AACTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAACGACCCCTCGTCACAGTAAG 2273

Qy   1660 TATTGAAGGACAGCCTGTAGAAGTATTACTGGATACAGGGGCTGATGATTCTATTGTAAC 1719
Db   2274 GATCGGGGGGCAACTCAAGGAAGCGCTGCTCGATACAGGAGCAGATGATACAGTATTAGA 2333

Qy   1720 AGGAATAGAGTTAGGTCCACATTATACCCCAAAAATAGTAGGAGGAATAGGAGGTTTTAT 1779
Db   2334 AGAAATGAGTTTGCCAGGAAGATGGAAACCAAAAATGATAGGGGGGATCGGGGGCTTCAT 2393

Qy   1780 TAATACTAAAGAATACAAAAATGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGAC 1839
Db   2394 CAAGGTGAGGCAGTACGACCAGATACTCATAGAAATCTGTGGACATAAAGCTATAGGTAC 2453

Qy   1840 AATCATGACAGGGGACACCCCGATTAACATTTTTGGTAGAAATTTGCTAAC 1890
Db   2454 AGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAATCTGTTGAC 2504
```

```
RESULT 12
US-09-968-355-24
; Sequence 24, Application US/09968355
; Patent No. US20020094523A1
; GENERAL INFORMATION:
;  APPLICANT: Sakalian, Michael
;  APPLICANT:  Hunter, Eric
;  TITLE OF INVENTION: Chimeric Retroviral Gag Genes and Screening Assays
;  FILE REFERENCE: UAB-100XC1
;  CURRENT APPLICATION NUMBER: US/09/968,355
;  CURRENT FILING DATE:  2001-09-28
;  PRIOR APPLICATION NUMBER: 60/236,273
;  PRIOR FILING DATE: 2000-09-28
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24
;   LENGTH: 4516
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Plasmid pDAB72 (FS-)
US-09-968-355-24

  Query Match             23.5%;  Score 450.4;  DB 10;  Length 4516;
  Best Local Similarity   56.4%;  Pred. No. 5.3e-111;
  Matches  972;  Conservative    0;  Mismatches  701;  Indels   51;  Gaps    5;

Qy    104 GATAGAGTGGGAGATGGGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATT 163
Db    259 GAAAGCTTAAACCATGGGTGCGAGAGCGTCAGTATTAAGCGGGGGAGAATTAGATCGATG 318

Qy    164 AGAAAAAATTAGGCTACGACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATG 223
Db    319 GGAAAAAATTCGGTTAAGGCCAGGGGGAAAGAAAAAATATAAATTAAAACATATAGTATG 378

Qy    224 GGCAGCAAATGAATTAGATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGG 283
Db    379 GGCAAGCAGGGAGCTAGAACGATTCGCAGTTAATCCTGGCCTGTTAGAAACATCAGAAGG 438

Qy    284 ATGTCAAAAAATACTTTCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAA 343
Db    439 CTGTAGACAAATACTGGGACAGCTACAACCATCCCTTCAGACAGGATCAGAAGAACTTAG 498

Qy    344 AAGCCTTTATAATACTGTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACA 403
Db    499 ATCATTATATAATACAGTAGCAACCCTCTATTGTGTGCATCAAAGGATAGAGATAAAAGA 558

Qy    404 CACTGAGGAAGCAAAACAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGA 463
Db    559 CACCAAGGAAGC------TTTAGACAAGATAGAGGAAGAGCAAAACAAAAGTAAGAAAAA 612

Qy    464 AACTATGCCAAAAACAAGTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCC 523
Db    613 AGCACAGCAAGCAGCAGCTGACACAGGACACAGCAATCAGGTCAGCCAAAATTACCCTAT 672

Qy    524 AGTACAACAAATAGGTGGTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGC 583
Db    673 AGTGCAGAACATCCAGGGGCAAATGGTACATCAGGCCATATCACCTAGAACTTTAAATGC 732

Qy    584 CTGGGTAAAATTGATAGAGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGC 643
Db    733 ATGGGTAAAAGTAGTAGAAGAGAAGGCTTTCAGCCCAGAAGTGATACCCATGTTTTCAGC 792

Qy    644 ACTGTCAGAAGGTTGCACCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCA 703
Db    793 ATTATCAGAAGGAGCCACCCCACAAGATTTAAACACCATGCTAAACACAGTGGGGGGACA 852

Qy    704 TCAAGCGGCTATGCAGATTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTT 763
Db    853 TCAAGCAGCCATGCAAATGTTAAAAGAGACCATCAATGAGGAAGCTGCAGAATGGGATAG 912

Qy    764 GCAGCACCCACAACCAG------CTCCACAACAAGGACAACTTAGGGAGCCGTCAGGATC 817
Db    913 AGTACATCCAGTGCATGCAGGGCCTATTGCACCAGGCCAGATGAGAGAACCAAGGGGAAG 972

Qy    818 AGATATTGCAGGAACAACTAGTTCAGTAGATGAACAAATCCAGTGGATGTACAGACAACA 877
Db    973 TGACATAGCAGGAACTACTAGTACCCTTCAGGAACAAATAGGATGGATG---ACAAATAA 1029

Qy    878 GAACCCCATACCAGTAGGCAACATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATG 937
Db   1030 TCCACCTATCCCAGTAGGAGAAATTTATAAAAGATGGATAATCCTGGGATTAAATAAAAT 1089

Qy    938 TGTCAGAATGTATAACCCAACAAACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATT 997
Db   1090 AGTAAGAATGTATAGCCCTACCAGCATTCTGGACATAAGACAAGGACCAAAAGAACCTTT 1149

Qy    998 TCAGAGCTATGTAGACAGGTTCTACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGT 1057
Db   1150 TAGAGACTATGTAGACCGGTTCTATAAAACTCTAAGAGCCGAGCAAGCTTCACAGGAGGT 1209

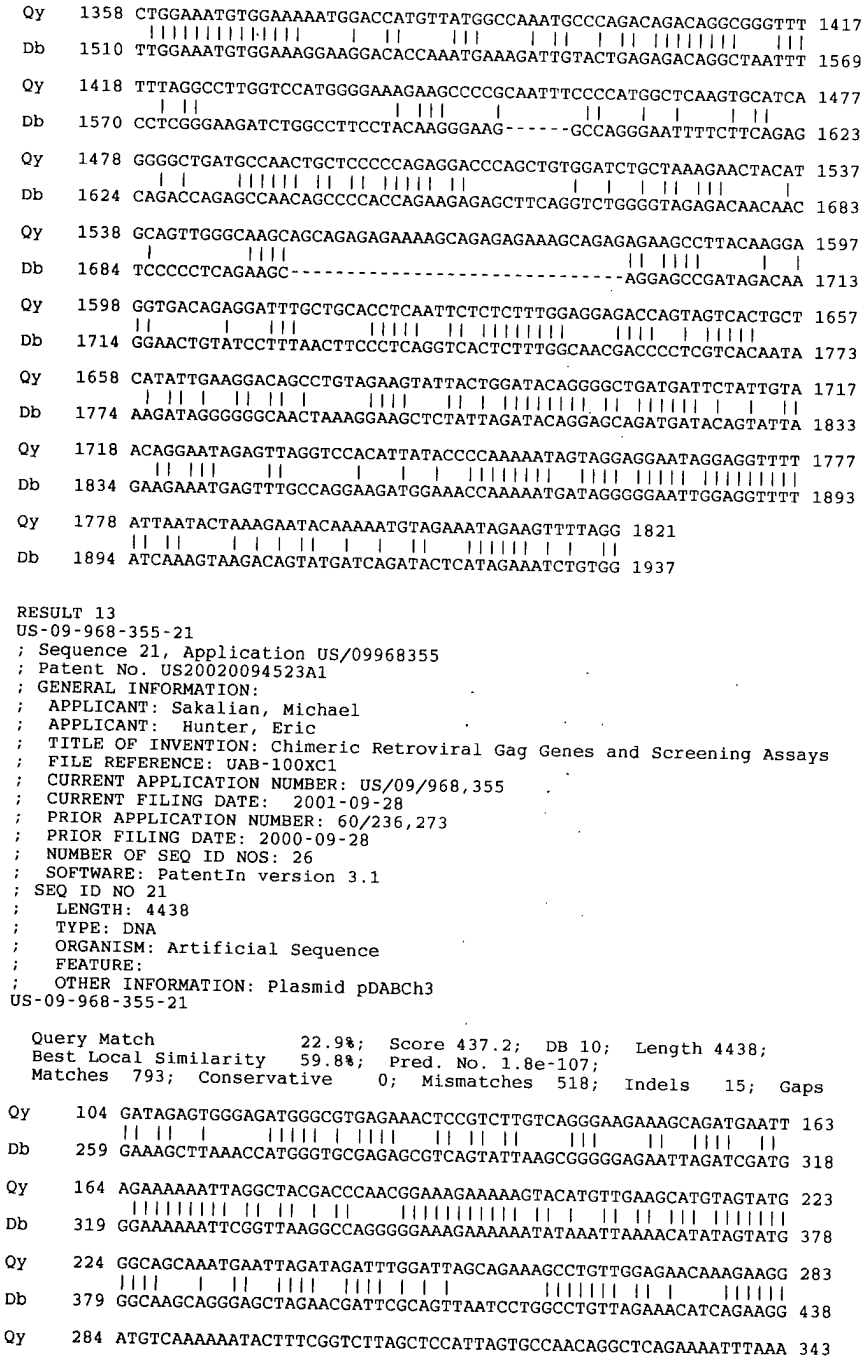
Qy   1058 AAAGAATTGGATGACTCAAACACTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGT 1117
Db   1210 AAAAAATTGGATGACAGAAACCTTGTTGGTCCAAAATGCGAACCCAGATTGTAAGACTAT 1269

Qy   1118 GCTGAAGGGGCTGGGTGTGAATCCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGT 1177
Db   1270 TTTAAAAGCATTGGGACCAGCGGCTACCTTAGAAGAAATGATGACAGCATGTCAGGGAGT 1329

Qy   1178 AGGGGGGCCGGGACAGAAGGCTAGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACC 1237
Db   1330 AGGAGGACCCGGCCATAAGGCAAGAGTTTTGGCTGAAGCAATGAGCCAAGTAACAAATAC 1389

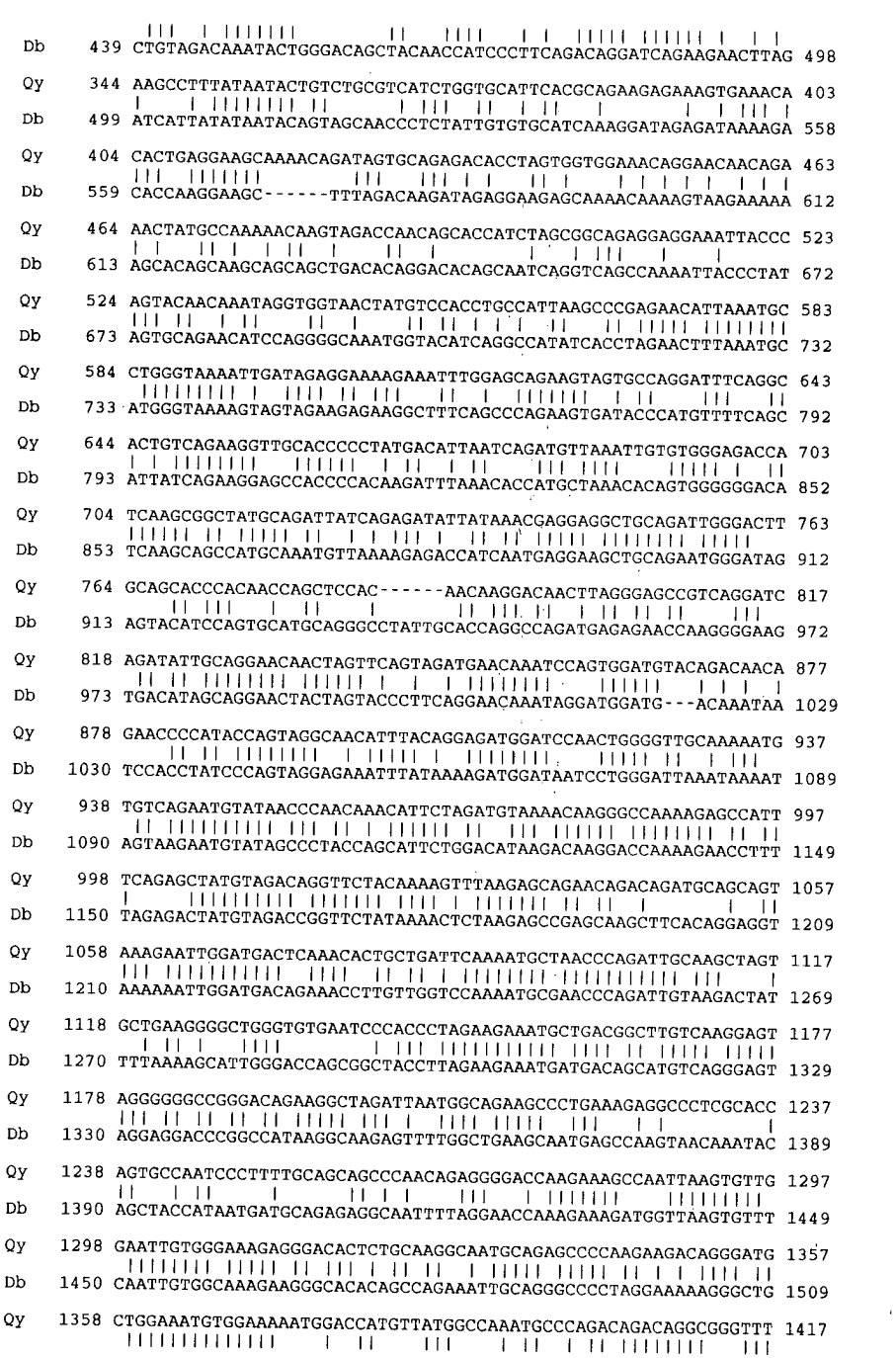
Qy   1238 AGTGCCAATCCCTTTTGCAGCAGCCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTG 1297
Db   1390 AGCTACCATAATGATGCAGAGAGGCAATTTTAGGAACCAAAGAAAGATGGTTAAGTGTTT 1449

Qy   1298 GAATTGTGGGAAAGAGGGACACTCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATG 1357
Db   1450 CAATTGTGGCAAAGAAGGGCACACAGCCAGAAATTGCAGGGCCCCTAGGAAAAAGGGCTG 1509
```



RESULT 13
US-09-968-355-21
; Sequence 21, Application US/09968355
; Patent No. US20020094523A1
; GENERAL INFORMATION:
; APPLICANT: Sakalian, Michael
; APPLICANT: Hunter, Eric
; TITLE OF INVENTION: Chimeric Retroviral Gag Genes and Screening Assays
; FILE REFERENCE: UAB-100XC1
; CURRENT APPLICATION NUMBER: US/09/968,355
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: 60/236,273
; PRIOR FILING DATE: 2000-09-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21
; LENGTH: 4438
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Plasmid pDABCh3
US-09-968-355-21

Query Match 22.9%; Score 437.2; DB 10; Length 4438;
Best Local Similarity 59.8%; Pred. No. 1.8e-107;
Matches 793; Conservative 0; Mismatches 518; Indels 15; Gaps 3;



```
Db   1510 TTGGAAATGTGGAAAGGAAGGACACCAAATGAAAGATTGTACTGAGAGACAGGCTAATTT 1569

Qy   1418 TTTAGG 1423
             | ||
Db   1570 CCTCGG 1575


RESULT 14
US-09-968-355-15
; Sequence 15, Application US/09968355
; Patent No. US20020094523A1
; GENERAL INFORMATION:
;  APPLICANT: Sakalian, Michael
;  APPLICANT:  Hunter, Eric
;  TITLE OF INVENTION: Chimeric Retroviral Gag Genes and Screening Assays
;  FILE REFERENCE: UAB-100XC1
;  CURRENT APPLICATION NUMBER: US/09/968,355
;  CURRENT FILING DATE:  2001-09-28
;  PRIOR APPLICATION NUMBER: 60/236,273
;  PRIOR FILING DATE: 2000-09-28
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15
;   LENGTH: 4594
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Plasmid pDABCh4
US-09-968-355-15

  Query Match             22.9%;  Score 437.2;  DB 10;  Length 4594;
  Best Local Similarity   59.8%;  Pred. No. 1.9e-107;
  Matches  793;  Conservative    0;  Mismatches  518;  Indels   15;  Gaps    3;

Qy    104 GATAGAGTGGGAGATGGGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATT 163
          || ||  |     ||||| | ||||   || || ||  ||||    ||  |||| ||
Db    259 GAAAGCTTAAACCATGGGTGCGAGAGCGTCAGTATTAAGCGGGGGAGAATTAGATCGATG 318

Qy    164 AGAAAAAATTAGGCTACGACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATG 223
          |||||||||| || || | ||    |||||||||||| || |  || || ||| |||||||
Db    319 GGAAAAAATTCGGTTAAGGCCAGGGGGAAAGAAAAAATATAAATTAAAACATATAGTATG 378

Qy    224 GGCAGCAAATGAATTAGATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGG 283
          ||||   |  ||  |||| |||| | | |        |||||| || |    ||||||
Db    379 GGCAAGCAGGGAGCTAGAACGATTCGCAGTTAATCCTGGCCTGTTAGAAACATCAGAAGG 438

Qy    284 ATGTCAAAAAATACTTTCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAA 343
          |||  | |||||||        ||   ||||   | |  ||||| |||||| |  | |
Db    439 CTGTAGACAAATACTGGGACAGCTACAACCATCCCTTCAGACAGGATCAGAAGAACTTAG 498

Qy    344 AAGCCTTTATAATACTGTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACA 403
          |    | |||||||| ||       | |||  ||  | ||   |       |  | ||| |
Db    499 ATCATTATATAATACAGTAGCAACCCTCTATTGTGTGCATCAAAGGATAGAGATAAAAGA 558

Qy    404 CACTGAGGAAGCAAAACAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGA 463
          |||  |||||||        |||   ||| | |   || |    | | | | |  | |
Db    559 CACCAAGGAAGC------TTTAGACAAGATAGAGGAAGAGCAAAACAAAAGTAAGAAAAA 612

Qy    464 AACTATGCCAAAAACAAGTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCC 523
          | |   || |  | ||  |    ||  |        |   | |||   |    |
Db    613 AGCACAGCAAGCAGCAGCTGACACAGGACACAGCAATCAGGTCAGCCAAAATTACCCTAT 672

Qy    524 AGTACAACAAATAGGTGGTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGC 583
          ||| ||  | ||    ||  |    || || | | |  ||   || ||||| ||||||||
Db    673 AGTGCAGAACATCCAGGGGCAAATGGTACATCAGGCCATATCACCTAGAACTTTAAATGC 732

Qy    584 CTGGGTAAAATTGATAGAGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGC 643
          |||||||||  |  |||| || |||   ||  |  |||||||  | ||    |||   ||
Db    733 ATGGGTAAAAGTAGTAGAAGAGAAGGCTTTCAGCCCAGAAGTGATACCCATGTTTTCAGC 792

Qy    644 ACTGTCAGAAGGTTGCACCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCA 703
          | | ||||||||   ||||||  | ||  | ||    ||| ||||    |||||  |  ||
Db    793 ATTATCAGAAGGAGCCACCCCACAAGATTTAAACACCATGCTAAACACAGTGGGGGGACA 852

Qy    704 TCAAGCGGCTATGCAGATTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTT 763
          |||||| || ||||| ||  | | ||| |  || || ||||| |||||||| |||||
Db    853 TCAAGCAGCCATGCAAATGTTAAAAGAGACCATCAATGAGGAAGCTGCAGAATGGGATAG 912

Qy    764 GCAGCACCCACAACCAGCTCCAC------AACAAGGACAACTTAGGGAGCCGTCAGGATC 817
             || |||    |  ||    |       || ||| ||  | || || ||    |||
Db    913 AGTACATCCAGTGCATGCAGGGCCTATTGCACCAGGCCAGATGAGAGAACCAAGGGGAAG 972

Qy    818 AGATATTGCAGGAACAACTAGTTCAGTAGATGAACAAATCCAGTGGATGTACAGACAACA 877
          || || |||||||| |||||| |  |  | ||||||||     ||||||   | | ||  |
Db    973 TGACATAGCAGGAACTACTAGTACCCTTCAGGAACAAATAGGATGGATG---ACAAATAA 1029

Qy    878 GAACCCCATACCAGTAGGCAACATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATG 937
             || || ||||||||  | |||||  |  ||||||||     |||||  ||  | |||
Db   1030 TCCACCTATCCCAGTAGGAGAAATTTATAAAAGATGGATAATCCTGGGATTAAATAAAAT 1089

Qy    938 TGTCAGAATGTATAACCCAACAAACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATT 997
          || |||||||||| ||| || | |||||| ||  ||| |||||| |||||||| || ||
Db   1090 AGTAAGAATGTATAGCCCTACCAGCATTCTGGACATAAGACAAGGACCAAAAGAACCTTT 1149

Qy    998 TCAGAGCTATGTAGACAGGTTCTACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGT 1057
          |     |||||||||| |||||| |||| | |||||| || ||  |       | | ||
Db   1150 TAGAGACTATGTAGACCGGTTCTATAAAACTCTAAGAGCCGAGCAAGCTTCACAGGAGGT 1209

Qy    1058 AAAGAATTGGATGACTCAAACACTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGT 1117
           ||| ||||||||||  |||| || || | |||||||| ||||||||| ||| | |
Db    1210 AAAAAATTGGATGACAGAAACCTTGTTGGTCCAAAATGCGAACCCAGATTGTAAGACTAT 1269

Qy    1118 GCTGAAGGGGCTGGGTGTGAATCCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGT 1177
            | || |  ||||        | ||| |||||||||| |||| || ||||| ||||
Db    1270 TTTAAAAGCATTGGGACCAGCGGCTACCTTAGAAGAAATGATGACAGCATGTCAGGGAGT 1329

Qy    1178 AGGGGGGCCGGGACAGAAGGCTAGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACC 1237
           ||| || || || || ||||| ||| |  |||| |||||  |||   | |         |
Db    1330 AGGAGGACCCGGCCATAAGGCAAGAGTTTTGGCTGAAGCAATGAGCCAAGTAACAAATAC 1389

Qy    1238 AGTGCCAATCCCTTTTGCAGCAGCCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTG 1297
           ||   | ||      |     || | |   |||   | |||||||    |||||||||
Db    1390 AGCTACCATAATGATGCAGAGAGGCAATTTTAGGAACCAAAGAAAGATGGTTAAGTGTTT 1449

Qy    1298 GAATTGTGGGAAAGAGGGACACTCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATG 1357
           |||||||| ||||| || ||| | || ||  | ||||| ||||| || | | |||| ||
Db    1450 CAATTGTGGCAAAGAAGGGCACACAGCCAGAAATTGCAGGGCCCCTAGGAAAAAGGGCTG 1509

Qy    1358 CTGGAAATGTGGAAAAATGGACCATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTT 1417
           | ||||||||||||||    |  ||    |||    | ||  | || |||||||   |||
Db    1510 TTGGAAATGTGGAAAGGAAGGACACCAAATGAAAGATTGTACTGAGAGACAGGCTAATTT 1569

Qy    1418 TTTAGG 1423
             | ||
Db    1570 CCTCGG 1575


RESULT 15
US-09-968-355-25
; Sequence 25, Application US/09968355
; Patent No. US20020094523A1
; GENERAL INFORMATION:
;  APPLICANT: Sakalian, Michael
;  APPLICANT:  Hunter, Eric
;  TITLE OF INVENTION: Chimeric Retroviral Gag Genes and Screening Assays
;  FILE REFERENCE: UAB-100XC1
;  CURRENT APPLICATION NUMBER: US/09/968,355
;  CURRENT FILING DATE:  2001-09-28
;  PRIOR APPLICATION NUMBER: 60/236,273
;  PRIOR FILING DATE: 2000-09-28
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25
;   LENGTH: 1503
;   TYPE: DNA
;   ORGANISM: Human immunodeficiency virus
```

```
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(1503)
;   OTHER INFORMATION: HIV Gag (FS-) DNA sequence
US-09-968-355-25

  Query Match             22.8%;  Score 437;  DB 10;  Length 1503;
  Best Local Similarity   60.0%;  Pred. No. 1.1e-107;
  Matches  788;  Conservative    0;  Mismatches  510;  Indels   15;  Gaps    3;

Qy    117 ATGGGCGTGAGAAACTCCGTCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGG 176
          |||||  | ||||   || || ||    |||    ||  ||||  ||  |||||||||  ||
Db      1 ATGGGTGCGAGAGCGTCAGTATTAAGCGGGGGAGAATTAGATCGATGGGAAAAAATTCGG 60

Qy    177 CTACGACCCAACGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAA 236
           || | ||    ||||||||||| || |  || || ||| ||||||||||    |  ||
Db     61 TTAAGGCCAGGGGGAAAGAAAAAATATAAATTAAAACATATAGTATGGGCAAGCAGGGAG 120

Qy    237 TTAGATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATA 296
          ||||  |||| | | |        ||||||| || |    |||||| |||  | |||||
Db    121 CTAGAACGATTCGCAGTTAATCCTGGCCTGTTAGAAACATCAGAAGGCTGTAGACAAATA 180

Qy    297 CTTTCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAAT 356
          ||        ||   ||||   | |  ||||| |||||| |  | | |    | ||||||
Db    181 CTGGGACAGCTACAACCATCCCTTCAGACAGGATCAGAAGAACTTAGATCATTATATAAT 240

Qy    357 ACTGTCTGCGTCATCTGGTGCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCA 416
          || ||      | |||  ||  | ||   |       |  | ||| ||||  |||||||
Db    241 ACAGTAGCAACCCTCTATTGTGTGCATCAAAGGATAGAGATAAAAGACACCAAGGAAGC- 299

Qy    417 AAACAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAA 476
                |||   ||| | |   || |    | | | | |  | | || |   || | |
Db    300 -----TTTAGACAAGATAGAGGAAGAGCAAAACAAAAGTAAGAAAAAAGCACAGCAAGCA 354

Qy    477 ACAAGTAGACCAACAGCACCATCTAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATA 536
          ||  |    ||  |        |   | |||   |    |.      ||| ||  | ||
Db    355 GCAGCTGACACAGGACACAGCAATCAGGTCAGCCAAAATTACCCTATAGTGCAGAACATC 414

Qy    537 GGTGGTAACTATGTCCACCTGCCATTAAGCCCGAGAACATTAAATGCCTGGGTAAAATTG 596
             ||  |    || || | | |  ||   || ||||| |||||||| |||||||| |
Db    415 CAGGGGCAAATGGTACATCAGGCCATATCACCTAGAACTTTAAATGCATGGGTAAAAGTA 474

Qy    597 ATAGAGGAAAAGAAATTTGGAGCAGAAGTAGTGCCAGGATTTCAGGCACTGTCAGAAGGT 656
          |||| || |||   ||  |  |||||||  | ||    |||   ||| | ||||||||
Db    475 GTAGAAGAGAAGGCTTTCAGCCCAGAAGTGATACCCATGTTTTCAGCATTATCAGAAGGA 534

Qy    657 TGCACCCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCAAGCGGCTATG 716
            ||||||  | ||  | ||    ||| ||||    ||||| |  |||||||| || |||
Db    535 GCCACCCCACAAGATTTAAACACCATGCTAAACACAGTGGGGGGACATCAAGCAGCCATG 594

Qy    717 CAGATTATCAGAGATATTATAAACGAGGAGGCTGCAGATTGGGACTTGCAGCACCCACAA 776
          || ||  | | ||| |  || || ||||| |||||||| |||||        || |||
Db    595 CAAATGTTAAAAGAGACCATCAATGAGGAAGCTGCAGAATGGGATAGAGTACATCCAGTG 654

Qy    777 CCAGCTCCAC------AACAAGGACAACTTAGGGAGCCGTCAGGATCAGATATTGCAGGA 830
          |  ||    |       || ||| ||  | || || ||    |||   || || ||||||
Db    655 CATGCAGGGCCTATTGCACCAGGCCAGATGAGAGAACCAAGGGGAAGTGACATAGCAGGA 714

Qy    831 ACAACTAGTTCAGTAGATGAACAAATCCAGTGGATGTACAGACAACAGAACCCCATACCA 890
          || |||||| |  |  | ||||||||    ||||||   | | |  |    || || |||
Db    715 ACTACTAGTACCCTTCAGGAACAAATAGGATGGATG---ACAAATAATCCACCTATCCCA 771

Qy    891 GTAGGCAACATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAATGTAT 950
          |||||  | ||||| |  ||||||||    ||||| ||  | |||   || |||||||||
Db    772 GTAGGAGAAATTTATAAAAGATGGATAATCCTGGGATTAAATAAAATAGTAAGAATGTAT 831

Qy    951 AACCCAACAAACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTA 1010
          | |||  || | |||||| ||  ||| |||||| |||||||| || |||     |||||||
Db    832 AGCCCTACCAGCATTCTGGACATAAGACAAGGACCAAAAGAACCTTTTAGAGACTATGTA 891

Qy   1011 GACAGGTTCTACAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATG 1070
          ||| ||||||| |||| | ||||||| || ||  |       |  ||||| |||||||||
Db    892 GACCGGTTCTATAAAACTCTAAGAGCCGAGCAAGCTTCACAGGAGGTAAAAAATTGGATG 951

Qy   1071 ACTCAAACACTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTG 1130
          ||  ||||  || || | |||||||| |||||||||| |||    |  | || |   ||
Db    952 ACAGAAACCTTGTTGGTCCAAAATGCGAACCCAGATTGTAAGACTATTTTAAAAGCATTG 1011

Qy   1131 GGTGTGAATCCCACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGA 1190
          ||         | ||| |||||||||| |||| || ||||| ||||||| || || ||
Db    1012 GGACCAGCGGCTACCTTAGAAGAAATGATGACAGCATGTCAGGGAGTAGGAGGACCCGGC 1071

Qy   1191 CAGAAGGCTAGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGTGCCAATCCCT 1250
          || ||||| ||| |  |||| |||||  |||   | |          |||   | ||
Db    1072 CATAAGGCAAGAGTTTTGGCTGAAGCAATGAGCCAAGTAACAAATACAGCTACCATAATG 1131

Qy   1251 TTTGCAGCAGCCCAACAGAGGGGACCAAGAAAGCCAATTAAGTGTTGGAATTGTGGGAAA 1310
            |      || | |    |||   | |||||||  ' |||||||||  |||||||| |||
Db    1132 ATGCAGAGAGGCAATTTTAGGAACCAAAGAAAGATGGTTAAGTGTTTCAATTGTGGCAAA 1191

Qy   1311 GAGGGACACTCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGA 1370
          || || ||| | || ||  | ||||| ||||| || | | |||| || ||||||||||||
Db    1192 GAAGGGCACACAGCCAGAAATTGCAGGGCCCCTAGGAAAAAGGGCTGTTGGAAATGTGGA 1251

Qy   1371 AAAATGGACCATGTTATGGCCAAATGCCCAGACAGACAGGCGGGTTTTTTAGG 1423
          ||    |  ||   |||    | ||  | || ||||||||   |||  | ||
Db    1252 AAGGAAGGACACCAAATGAAAGATTGTACTGAGAGACAGGCTAATTTCCTCGG 1304


Search completed: November 16, 2002, 13:10:17
Job time : 83.6447 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 16, 2002, 07:32:18 ; Search time 1772.71 Seconds
(without alignments)
17477:179 Million cell updates/sec

Title: M33262_COPY_1193_3105
Perfect score: 1913
Sequence: 1 GGTACCAGACGGCGTGAGGA.....TCTGGGGATGTCTCTAAATT 1913

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	62.2	3.3	997	17	CNS005TE	AL060767 Drosophil
c	2	61.2	3.2	987	17	CNS00418	AL066537 Drosophil
c	3	58.2	3.0	450	9	AA677404	AA677404 zj62g08.s
	4	58.2	3.0	664	17	AQ196120	AQ196120 RPCI11-44
	5	57.4	3.0	860	17	AQ742940	AQ742940 HS_5386_B
	6	56.2	2.9	476	17	AQ032180	AQ032180 HS_2193_B
	7	53.8	2.8	850	17	AQ780780	AQ780780 HS_3104_B
	8	53.2	2.8	345	17	AQ070561	AQ070561 HS_2268_B
c	9	51.8	2.7	436	9	AI124782	AI124782 am56b01.x
c	10	51.8	2.7	470	9	AA831618	AA831618 oc80d10.s
	11	51.8	2.7	553	17	AQ037620	AQ037620 CIT-HSP-2
c	12	51.8	2.7	571	10	AW629027	AW629027 h146g08.x
	13	50.6	2.6	1300	13	BM468018	BM468018 AGENCOURT
c	14	50.2	2.6	496	14	R93353	R93353 yt69a11.r1
	15	49.2	2.6	491	17	B56450	B56450 CIT-HSP-200
	16	48.4	2.5	599	17	CNS008XP	AL052566 Drosophil
	17	48.4	2.5	972	17	CNS00JAT	AL075943 Drosophil
c	18	48.4	2.5	983	17	CNS014HQ	AL104168 Drosophil
	19	48.2	2.5	1101	17	CNS017YG	AL108658 Drosophil
	20	47.8	2.5	1107	9	AL560930	AL560930 AL560930
	21	47.2	2.5	547	9	AL601859	AL601859 DKFZp313L
c	22	47	2.5	974	17	CNS00ITT	AL075432 Drosophil
c	23	46.6	2.4	879	17	CNS01JRG	AL147405 Anopheles
c	24	46.4	2.4	838	17	CNS03W9J	AL263440 Tetraodon
c	25	46.4	2.4	1204	17	CNS016E2	AL106628 Drosophil
c	26	46	2.4	1096	17	B09949	B09949 T1F7-Sp6 TA
	27	46	2.4	1101	17	CNS0006J	AL062049 Drosophil
c	28	45.8	2.4	1101	17	CNS017QV	AL108385 Drosophil
c	29	45.4	2.4	824	17	CNS009UF	AL053946 Drosophil
	30	45.2	2.4	647	13	BJ393377	BJ393377 BJ393377
	31	45	2.4	984	12	BG844058	BG844058 1024004H1
c	32	45	2.4	1101	17	CNS00DG1	AL069971 Drosophil
c	33	44.8	2.3	366	12	BG001163	BG001163 QV1-GN020
	34	44.8	2.3	875	17	AZ529564	AZ529564 ENTDC90TF
	35	44.8	2.3	889	17	AZ530247	AZ530247 ENTCJ40TF
c	36	44.8	2.3	890	17	AZ675153	AZ675153 ENTKW94TF
	37	44.8	2.3	918	17	BH139392	BH139392 ENTNK40TR
	38	44.8	2.3	932	17	AZ681268	AZ681268 ENTMB47TR
	39	44.8	2.3	932	17	BH139815	BH139815 ENTPA66TF
	40	44.8	2.3	933	17	BH159320	BH159320 ENTSJ63TR
	41	44.8	2.3	970	17	BH132580	BH132580 ENTNZ22TF
c	42	44.8	2.3	1398	13	BM416038	BM416038 OP21121 M
	43	44.6	2.3	564	10	AW668618	AW668618 ZA51 DDRT
	44	44.6	2.3	913	17	CNS04D8L	AL285438 Tetraodon
	45	44.4	2.3	562	13	BM168066	BM168066 EST570589

ALIGNMENTS

RESULT 1
CNS005TE/c
LOCUS CNS005TE 997 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR12K22 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL060767
VERSION AL060767.1 GI:4943573
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 997)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial

```
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .997
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR12K22"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
BASE COUNT       89 a     99 c     13 g    258 t    538 others
ORIGIN

  Query Match             3.3%;  Score 62.2;  DB 17;  Length 997;
  Best Local Similarity   17.7%;  Pred. No. 0.00083;
  Matches   92;  Conservative  189;  Mismatches  238;  Indels    0;  Gaps    0;

Qy      20 AGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGCAACACAAAAAAGAAATAGCTGT 79
           :: :::|||: :|:|::     :|  | :|  |:: : :: : |:| |:.|:| |:  |
Db     996 RRARRRAGARARARARRGARRRGRRGARGARARRARRRRARGARAGARAARAGARRAGA 937

Qy      80 CTTTTATCCAGGAAGGGGTAATAAGATAGAGTGGGAGATGGGCGTGAGAAACTCCGTCTT 139
             |   :: :::: | :| :||: ::|| |:|:|  :      |::|:
Db     936 GARARAAARRRARRRRAGRRAGRAGRGRSAGGGRGRGGRRAAAARAARDARARARARARA 877

Qy     140 GTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCTACGACCCAACGGAAAGAAAAA 199
           :  |: :||:|:   ::: :::   :| :::|  |||    :     : : :|:::| ||
Db     876 RRRARAKAARARGARRRRGRRRRRGRAGRRRARGAGGRAGRRGRGRGRRRARARRRAGAA 817

Qy     200 GTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTAGATAGATTTGGATTAGCAGA 259
           : : : :  :||: : : ::   ::: |: |:  :::  ::: :|    |::  |  : :
Db     816 RARARGRRARAARARARRRRAGRRRRRARGARGRRRRGGRRRGRGGRRRGRRRRAARRAR 757

Qy     260 AAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTTTCGGTCTTAGCTCCATTAGT 319
           |:          :::: ::|||| :: :  :::::::       ::    |:    |
Db     756 ARARAGARAAARRRRARRAGAAARRARRRRRRRRRRGAGARARRARRRARRAARAGAGAA 697

Qy     320 GCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACTGTCTGCGTCATCTGGTGCAT 379
           :  :  :|:   |::: :   ::: ||     : :: :  : : : :  :       : |
Db     696 RRRRGRRGRGAGARRRGRGRGRRRGAGARRRRRGRRMRTRRARRRRRARRGAGAARRRAR 637

Qy     380 TCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAACAGATAGTGCAGAGACACCT 439
           : : :|::|::: : : :: :    :::|::   |:: ::: :  : ::|:| |
Db     636 GARRRRRGRRGRRRRGRARGRRRRAGRGRRRGRRARGARRGRRRAGRRRRRRARAGAGAR 577

Qy     440 AGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAAACAAGTAGACCAACAGCACCATC 499
           :| :: |:|:: |:||| :: |::|:  : :  |:: | :|| : :  |  |: |  |
Db     576 RGRRRRGRARRAARGAARRRRARRARRRRRRAAARRGAGRAGARARGRAGRARRARRARR 517

Qy     500 TAGCGGCAGAGGAGGAAATTACCCAGTACAACAAATAGG 538
           :: ::   ::: |::    :   |: : || ||| |::
Db     516 RRRRRRGGARRRGGRRGGRRRAAGARRRRAARAAARARR 478


RESULT 2
CNS00418/c
LOCUS       CNS00418                 987 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR09C16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL066537
VERSION     AL066537.1  GI:4942778
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 987)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .987
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR09C16"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
BASE COUNT      238 a    162 c     17 g    177 t    393 others
ORIGIN

  Query Match             3.2%;  Score 61.2;  DB 17;  Length 987;
  Best Local Similarity   22.7%;  Pred. No. 0.0014;
  Matches  114;  Conservative  150;  Mismatches  238;  Indels    0;  Gaps    0;

Qy      16 GAGGAGCGGGAGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGCAACACAAAAAAGAAATAG 75
           :|: :: :: : || |::: |          | |:: ::  : :| | :|||:: |:| ::
Db     955 RARARRRRARAAGAARRRAGRAGGRAAARGAARRRRRAARARAAAARAAARRAARAARR 896

Qy      76 CTGTCTTTTATCCAGGAAGGGGTAATAAGATAGAGTGGGAGATGGGCGTGAGAAACTCCG 135
             :       :   |:: :| |: :: : :: ::|| :::::: |:| | |:|::     :
Db     895 AARRRAARRRAAGARRGRGAGRARRGRGRRGRRAGRRRRRRRRGRGRGAGRGRRGRRRRR 836

Qy     136 TCTTGTCAGGGAAGAAAGCAGATGAATTAGAAAAAATTAGGCTACGACCCAACGGAAAGA 195
             :  ::|:::|:::|  ::  |:  :  |:|::  |:|  : |:   :  |||||  :
Db     835 RGRRRAGRRGRRRGRRRGRGRRAAARRRRAGARARRRRARGRGRRGRRGARGRGGAAAAR 776

Qy     196 AAAAGTACATGTTGAAGCATGTAGTATGGGCAGCAAATGAATTAGATAGATTTGGATTAG 255
           |:|:: : | :  |:::| | | ||   :|| |  :||  ||  | | |:|   : |  ||
Db     775 ARARRARAAARRRGRRGGARGGAGAGARGGAAARRAAAAAAARAAAGARAAAARAAARAG 716

Qy     256 CAGAAAGCCTGTTGGAGAACAAAGAAGGATGTCAAAAAATACTTTCGGTCTTAGCTCCAT 315
            |:|::|   :  |  | | | :| ::|: |   : :         |     ::    :
Db     715 RARARRGARGRGRGAGGGAGAGRGGRRGRGGRGRGGRGARRRRGAGRRRRRGARGRG 656

Qy     316 TAGTGCCAACAGGCTCAGAAAATTTAAAAAGCCTTTATAATACTGTCTGCGTCATCTGGT 375
           :: |  |  :|:    |:| |   ::|: :     : |:    :   : : :   |
Db     655 GRRRGAGAGGRGRRGRGGRAGARRRRRARGRRAGAGRGARRGRGRRRGRRRARRAGAGAR 596

Qy     376 GCATTCACGCAGAAGAGAAAGTGAAACACACTGAGGAAGCAAAACAGATAGTGCAGAGAC 435
           : |   : :  ::||::|||:  |:    |  | |  :| :  : ::  :: |  | |:
Db     595 RGARRRRARRGRRAGRRAAARAAARGRGGARAGGGAGRGGRGGRGRRGGRRGGGGGGRG 536

Qy     436 ACCTAGTGGTGGAAACAGGAACAACAGAAACTATGCCAAAAACAAGTAGACCAACAGCAC 495
           :    | |: :::::  |::| :: : :::    :  ||::| :|  |    |:  : :
Db     535 RRGRGGRGRGRRRRRRGGRRAGRRARARRRRGGRRAGAARRAARAAGAAGAGARAGRGRR 476

Qy     496 CATCTAGCGGCAGAGGAGGAAA 517
           :   :| |   ||:|||::||
Db     475 RRRGRRGGGARGGARGAGRRAA 454


RESULT 3
AA677404/c
LOCUS       AA677404                 450 bp    mRNA    linear   EST 19-DEC-1997
```

```
DEFINITION  zj62g08.s1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA
            clone IMAGE:454910 3' similar to TR:G887448 G887448 GAG MRNA. ;,
            mRNA sequence.
ACCESSION   AA677404
VERSION     AA677404.1  GI:2657926
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
            ,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Possible reversed clone: similarity on wrong strand
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 444.
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:454910"
                     /clone_lib="Soares_fetal_liver_spleen_1NFLS_S1"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     This is a subtracted version of the original Soares fetal
                     liver spleen 1NFLS library.  1st strand cDNA was primed
                     with a Pac I - oligo(dT) primer [5'
                     AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      117 a    120 c     84 g    129 t
ORIGIN

  Query Match             3.0%;  Score 58.2;  DB 9;  Length 450;
  Best Local Similarity   51.3%;  Pred. No. 0.0056;
  Matches  135;  Conservative    0;  Mismatches  128;  Indels    0;  Gaps    0;

Qy     962 CATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCTA 1021
           | ||   |  |||||||| || |||||||| || |        | | |||   ||||  |
Db     334 CTTTAGTGCAGTAAAACAGGGACCAAAAGAACCGTACACTGATTGTATAGCTCGGTTACA 275

Qy    1022 CAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACACT 1081
             |   | | | |   |     ||||||| | || |   || ||    ||   ||   |
Db     274 GGAGTCTCTTAAAAAGGTGATCACAGATTCGGCTGCTCAGGATATAGTGCTGCAGTTATC 215

Qy    1082 GCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATCC 1141
               | | |||||||| || |||||| ||   |  |||       |  | | |||  |
Db     214 AGCTTTCGACAATGCTAATCCCGATTGCCAGGCTGCTCTGTGACCTATCAGAGGGAAAGC 155

Qy    1142 CACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCTAG 1201
           | |||  ||     | | ||| ||| | ||  |  | ||  |    || || || |
Db     154 ACACTTAGTTGATTATATCAAGGCCTGTGATGGTATCAGAGGTACTCTGCATAAAGCGAC 95

Qy    1202 ATTAATGGCAGAAGCCCTGAAAG 1224
             ||  ||||| ||||  ||  ||
```

```
Db      94 TTTGTTGGCACAAGCTATGGCAG 72


RESULT 4
AQ196120
LOCUS       AQ196120                664 bp    DNA     linear   GSS 20-APR-1999
DEFINITION  RPCI11-44M5.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-44M5, DNA
            sequence.
ACCESSION   AQ196120
VERSION     AQ196120.1  GI:3607732
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 664)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
  TITLE     Use of human BAC End Sequences for Sequence-Ready Map Building
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .664
                     /organism="Homo sapiens"
                     /db_xref="GDB:7516804"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-44M5"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      192 a    127 c    163 g    178 t      4 others
ORIGIN

  Query Match             3.0%;  Score 58.2;  DB 17;  Length 664;
  Best Local Similarity   51.3%;  Pred. No. 0.0063;
  Matches  135;  Conservative    0;  Mismatches  128;  Indels    0;  Gaps    0;

Qy     962 CATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCTA 1021
           | ||   |  |||||||| || |||||||| ||||    |   | | |||   |||  |
Db     278 CTTTAGTGCTATAAAACAGGGACCAAAAGAACCATACGTGGATTTTATAGCTCAGTTACA 337

Qy    1022 CAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACACT 1081
             |   | | | |           ||||| | || ||  || || |  ||   ||   | |
Db     338 GTAGTCTCTTAAAAAGATGATTGCAGATTCGGCTGTTCAGGATAGAGTGTTGCAGTTATT 397

Qy    1082 GCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATCC 1141
               | | |||||||| || |||||| ||   |  |||       |  | | |||  |
Db     398 AGCTTTCGACAATGCTAATCCCGATTGCCAGGCTGCTCTGCGACCTATCAGAGGGAAAGC 457

Qy    1142 CACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCTAG 1201
           |||  ||     | | ||| ||| | ||  | || ||       || |||||||
Db     458 ACATTTAGTTGATTATATCAAGGCCTGTGATGGTATCGGAGGTAATCTGCATAAGGCTAC 517

Qy    1202 ATTAATGGCAGAAGCCCTGAAAG 1224
             |  | ||| | ||  ||  ||
Db     518 TCTGCTAGCACAGGCAGTGGCAG 540
```

```
RESULT 5
AQ742940
LOCUS       AQ742940                 860 bp    DNA     linear   GSS 16-JUL-1999
DEFINITION  HS_5386_B2_C06_T7A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=962 Col=12 Row=F, DNA sequence.
ACCESSION   AQ742940
VERSION     AQ742940.1  GI:5520462
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 860)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 962  row: F  column: 12
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 860.
FEATURES             Location/Qualifiers
     source          1. .860
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=962 Col=12 Row=F"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT      242 a    168 c    207 g    240 t      3 others
ORIGIN

  Query Match             3.0%;  Score 57.4;  DB 17;  Length 860;
  Best Local Similarity   51.4%;  Pred. No. 0.01;
  Matches  133;  Conservative    0;  Mismatches  126;  Indels    0;  Gaps    0;

Qy     962 CATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCTA 1021
           | ||   | | |||||||| || |||||||| ||||     |  | | |||    |||  |
Db     504 CTTTAGTGCTATAAAACAGGGACCAAAAGAACCATACGTGGATTTTATAGCTCAGTTACA 563

Qy    1022 CAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACACT 1081
             |   | | | |          ||||| | || ||  || || |  ||   ||   | |
Db     564 GTAGTCTCTTAAAAAGATGATTGCAGATTCGGCTGTTCAGGATAGAGTGTTGCAGTTATT 623

Qy    1082 GCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATCC 1141
             |  | |||||||| || |||||| ||   | |||          |  | | |||  |
Db     624 AGCTTTCGACAATGCTAATCCCGATTGCCAGGCTGCTCTGCGACCTATCAGAGGGAAAGC 683

Qy    1142 CACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCTAG 1201
           |||  ||     | | ||| ||| | ||  | || ||       || |||||||
Db     684 ACATTTAGTTGATTATATCAAGGCCTGTGATGGTATTGGAGGTAATCTGCATAAGGCTAC 743

Qy    1202 ATTAATGGCAGAAGCCCTG 1220
```

```
           |  | |||  | ||  ||
Db     744 TCTGCTAGCACAGGCAATG 762


RESULT 6
AQ032180
LOCUS       AQ032180                 476 bp    DNA     linear   GSS 02-JUL-1998
DEFINITION  HS_2193_B2_C11_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2193 Col=22 Row=F, DNA sequence.
ACCESSION   AQ032180
VERSION     AQ032180.1  GI:3284368
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 476)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 2193  row: F  column: 22
            Class: BAC ends
            High quality sequence stop: 476.
FEATURES             Location/Qualifiers
     source          1. .476
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=2193 Col=22 Row=F"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      136 a     97 c    115 g    123 t      5 others
ORIGIN

  Query Match             2.9%;  Score 56.2;  DB 17;  Length 476;
  Best Local Similarity   50.6%;  Pred. No. 0.017;
  Matches  133;  Conservative    0;  Mismatches  130;  Indels    0;  Gaps    0;

Qy     962 CATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCTA 1021
           | ||   | | |||||||| || |||||| |||| |       | | |||   ||||  |
Db     193 CTTTAGTGCTATAAAACAGGGACCCAAAGAACCATATGTTGATTTTATAGCTCGGTTACA 252

Qy    1022 CAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACACT 1081
             |   | | | |          ||||| | || |   || ||    ||   ||   | |
Db     253 GGAGTCTCTTANAAAGGTGATTGCAGATTCGGCTGCTCAGCATATAGTGTTGCAGTTATT 312

Qy    1082 GCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATCC 1141
             |  | |||||||| ||  ||||| |||  |  |||          |  | |||  |
Db     313 AGCTTTCGACAATGCTAATCCCAATTGCCAGCCTGCTCTGTGACCTATCANAGGGAAAGC 372

Qy    1142 CACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCTAG 1201
           |||  ||     | | ||| ||| | ||  | || ||          || || |||||
Db     373 ACATTTAGTTGATTATATCAAGGCCTGTGATGGTATCGGAGGTAATCTGCATAAAGCTAG 432

Qy    1202 ATTAATGGCAGAAGCCCTGAAAG 1224
             |  | |||  | ||  ||  ||
Db     433 TCTGCTAGCACAGGCAGTGGCAG 455


RESULT 7
```

```
AQ780780
LOCUS       AQ780780                 850 bp    DNA     linear   GSS 02-AUG-1999
DEFINITION  HS_3104_B1_C03_T7C CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3104 Col=5 Row=F, DNA sequence.
ACCESSION   AQ780780
VERSION     AQ780780.1  GI:5683740
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 850)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 3104  row: F  column: 5
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 850.
FEATURES             Location/Qualifiers
     source          1. .850
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=3104 Col=5 Row=F"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      244 a    156 c    206 g    244 t
ORIGIN

  Query Match             2.8%;  Score 53.8;  DB 17;  Length 850;
  Best Local Similarity   49.1%;  Pred. No. 0.072;
  Matches  142;  Conservative    0;  Mismatches  147;  Indels    0;  Gaps    0;

Qy     929 GCAAAAATGTGTCAGAATGTATAACCCAACAAACATTCTAGATGTAAAACAAGGGCCAAA 988
           | ||||||   |     || | |||   ||      | |||   || || | || ||||
Db     455 GAAAAAATCACTATAGGTGGAGAACATTACCCTTCCTTTAGTGCTATAAAACGGACCAAG 514

Qy     989 AGAGCCATTTCAGAGCTATGTAGACAGGTTCTACAAAAGTTTAAGAGCAGAACAGACAGA 1048
           ||| |||| |      | | |||    |||  |  |   | | | |         ||||
Db     515 AGAACCATATGTTGATTTTATAGCTCAGTTACAGGAGTCTCTTAAAAAGATGATTGCAGA 574

Qy    1049 TGCAGCAGTAAAGAATTGGATGACTCAAACACTGCTGATTCAAAATGCTAACCCAGATTG 1108
           | | || |   ||  |    ||   ||   | |     || | |||||||  | |||||
Db     575 TTCGGCTGCTCAGGTTATAGTGTTGCAGTTATTAGCTTTTGACAATGCTAATGCCGATTG 634

Qy    1109 CAAGCTAGTGCTGAAGGGGCTGGGTGTGAATCCCACCCTAGAAGAAATGCTGACGGCTTG 1168
           | ||   |  ||| |     |  | | |||  |     |||  ||     | | ||| ||
Db     635 CCAGGCTGCTCTGCAACCTGTCAGAGGGAAAGCACATTTAGTTGATTACATCAAGGCCTG 694

Qy    1169 TCAAGGAGTAGGGGGGCCGGGACAGAAGGCTAGATTAATGGCAGAAGCC 1217
           | | ||  |  | ||       || || ||||  ||  ||||| | |||
Db     695 TGATGGTATCAGAGGTAATCTGCATAAAGCTACTTTGTTGGCACAGGCC 743


RESULT 8
AQ070561
LOCUS       AQ070561                 345 bp    DNA     linear   GSS 05-AUG-1998
DEFINITION  HS_2268_B2_C02_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2268 Col=4 Row=F, DNA sequence.
ACCESSION   AQ070561
VERSION     AQ070561.1  GI:3388360
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 345)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 2268  row: F  column: 4
            Class: BAC ends
            High quality sequence stop: 345.
FEATURES             Location/Qualifiers
     source          1. .345
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=2268 Col=4 Row=F"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      102 a     60 c     87 g     96 t
ORIGIN

  Query Match             2.8%;  Score 53.2;  DB 17;  Length 345;
  Best Local Similarity   48.7%;  Pred. No. 0.075;
  Matches  145;  Conservative    0;  Mismatches  153;  Indels    0;  Gaps    0;

Qy     929 GCAAAAATGTGTCAGAATGTATAACCCAACAAACATTCTAGATGTAAAACAAGGGCCAAA 988
           | ||||||   |     || | |||   ||      | |||   || || | || ||||
Db      48 GAAAAAATCACTATAGGTGGAGAACATTACCCTTCCTTTAGTGCTATAAAACGGACCAAG 107

Qy     989 AGAGCCATTTCAGAGCTATGTAGACAGGTTCTACAAAAGTTTAAGAGCAGAACAGACAGA 1048
           ||| |||| |      | | |||    |||  |  |   | | | |         ||||
Db     108 AGAACCATATGTTGATTTTATAGCTCAGTTACAGGAGTCTCTTAAAAAGATGATTGCAGA 167

Qy    1049 TGCAGCAGTAAAGAATTGGATGACTCAAACACTGCTGATTCAAAATGCTAACCCAGATTG 1108
           | | || |   ||  |    ||   ||   | |     || | |||||||  | |||||
Db     168 TTCGGCTGCTCAGGTTATAGTGTTGCAGTTATTAGCTTTTGACAATGCTAATGCCGATTG 227

Qy    1109 CAAGCTAGTGCTGAAGGGGCTGGGTGTGAATCCCACCCTAGAAGAAATGCTGACGGCTTG 1168
           | ||   |  ||| |     |  | | |||  |     |||  ||     | | ||| ||
Db     228 CCAGGCTGCTCTGCAACCTGTCAGAGGGAAAGCACATTCAGTTGATTACATCAAGGCCTG 287

Qy    1169 TCAAGGAGTAGGGGGGCCGGGACAGAAGGCTAGATTAATGGCAGAAGCCCTGAAAGAG 1226
           | | ||  |  | ||       || || ||||  ||  ||||| | ||  ||  || |
Db     288 TGATGGTATTAGAGGTAATCTGCATAAAGCTACTTTGTTGGCACAGGCAATGGCAGGG 345


RESULT 9
AI124782/c
LOCUS       AI124782                 436 bp    mRNA    linear   EST 11-SEP-1998
DEFINITION  am56b01.x1 Johnston frontal cortex Homo sapiens cDNA clone
            IMAGE:1539529 3' similar to TR:P87891 P87891 GAG PROTEIN ;, mRNA
            sequence.
ACCESSION   AI124782
VERSION     AI124782.1  GI:3593296
KEYWORDS    EST.
```

```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 436)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
            ,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Possible reversed clone: similarity on wrong strand
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 424.
FEATURES             Location/Qualifiers
     source          1. .436
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1539529"
                     /clone_lib="Johnston frontal cortex"
                     /sex="male"
                     /tissue_type="pooled frontal lobe"
                     /dev_stage="adult"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: brain; Vector: Bluescript SK-; Site_1: EcoRI
                     ; Stanley Neuropathology Consortium (www.stanleylab.org)
                     brains S-58, S-65,  S-67, S-78. Random + oligo-dT primed
                     into EcoRI site of ZAP II Vector. Mass excised.  Avg
                     insert length 1.9kb. Custom library provided by Dr. Nancy
                     Johnston [(410) 614-3918,  nlj@welchlink.welch.jhu.edu].
                     "
BASE COUNT      110 a    114 c     79 g    133 t
ORIGIN

  Query Match             2.7%;  Score 51.8;  DB 9;  Length 436;
  Best Local Similarity   49.8%;  Pred. No. 0.17;
  Matches  131;  Conservative    0;  Mismatches  132;  Indels    0;  Gaps    0;

Qy     962 CATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCTA 1021
           | ||   |  |||||||| || |||||||| || |      | | |||    |||  |
Db     383 CTTTAGTGCAGTAAAACAGGGACCAAAAGAACCGTACGCAGATTTTATAGCTCAGTTACA 324

Qy    1022 CAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACACT 1081
           |   | | | |    |      ||||| | || |    || ||    ||   ||   |
Db     323 GGAGTCTCTTAAAAAGGTGATTGCAGATTCGGCTGCTCAGGATATAGTGCTGCAGTTATC 264

Qy    1082 GCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATCC 1141
           |  | |||||||| || |||||| ||    |  |||          |  | | |||  |
Db     263 AGCTTTCGACAATGCTAATCCCGATTGCCAGGCTGCTCTGTGACCTATCAGAGGGAACGC 204

Qy    1142 CACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCTAG 1201
           | |||  ||      | | ||| ||| | ||  |  | ||          || || ||||
Db     203 ACACTTAGTTGATTATATCAAGGCCTGTGACGGTATCAGAGGTAATATGCATAAAGCTAC 144

Qy    1202 ATTAATGGCAGAAGCCCTGAAAG 1224
           ||  ||| | | ||  ||  ||
Db     143 TTTGTTGGTACAGGCAATGGCAG 121


RESULT 10
AA831618/c
LOCUS       AA831618                 470 bp    mRNA    linear   EST 07-APR-1998
DEFINITION  oc80d10.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1356019 3'
            similar to TR:P87891 P87891 GAG PROTEIN ;, mRNA sequence.
ACCESSION   AA831618
VERSION     AA831618.1  GI:2904717
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 470)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 623    Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 464.
FEATURES             Location/Qualifiers
     source          1. .470
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1356019"
                     /clone_lib="NCI_CGAP_GCB1"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-3'
                     ]. Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      116 a    123 c     82 g    149 t
ORIGIN

  Query Match             2.7%;  Score 51.8;  DB 9;  Length 470;
  Best Local Similarity   49.8%;  Pred. No. 0.18;
  Matches  131;  Conservative    0;  Mismatches  132;  Indels    0;  Gaps    0;

Qy     962 CATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCTA 1021
           | ||   |  |||||||| || |||||||| || |      | | |||    |||  |
Db     390 CTTTAGTGCAGTAAAACAGGGACCAAAAGAACCGTACGCAGATTTTATAGCTCAGTTACA 331

Qy    1022 CAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACACT 1081
           |   | | | |    |      ||||| | || |    || ||    ||   ||   |
Db     330 GGAGTCTCTTAAAAAGGTGATTGCAGATTCGGCTGCTCAGGATATAGTGCTGCAGTTATC 271

Qy    1082 GCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATCC 1141
           |  | |||||||| || |||||| ||    |  |||          |  | | |||  |
Db     270 AGCTTTCGACAATGCTAATCCCGATTGCCAGGCTGCTCTGTGACCTATCAGAGGGAACGC 211

Qy    1142 CACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCTAG 1201
           | |||  ||      | | ||| ||| | ||  |  | ||          || || ||||
Db     210 ACACTTAGTTGATTATATCAAGGCCTGTGACGGTATCAGAGGTAATATGCATAAAGCTAC 151

Qy    1202 ATTAATGGCAGAAGCCCTGAAAG 1224
           ||  ||| | | ||  ||  ||
Db     150 TTTGTTGGTACAGGCAATGGCAG 128
```

```
RESULT 11
AQ037620
LOCUS       AQ037620                553 bp    DNA     linear   GSS 11-JUL-1998
DEFINITION  CIT-HSP-2337C22.TR CIT-HSP Homo sapiens genomic clone 2337C22, DNA
            sequence.
ACCESSION   AQ037620
VERSION     AQ037620.1  GI:3303452
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 553)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
            Simon,M. and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building (1998)
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: CIT-HSP-2337C22.TF
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .553
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="2337C22"
                     /clone_lib="CIT-HSP"
                     /sex="Male"
                     /cell_type="Sperm"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
BASE COUNT      165 a     98 c    131 g    159 t
ORIGIN

  Query Match             2.7%;  Score 51.8;  DB 17;  Length 553;
  Best Local Similarity   49.8%;  Pred. No. 0.18;
  Matches  131;  Conservative    0;  Mismatches  132;  Indels    0;  Gaps    0;

Qy    962 CATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCTA 1021
          | ||   | | ||||||| || |||| |||  |||    |  ||||   ||||  |
Db    202 CTTTAGTGCTATAAAACAGGGACCAAGAGAAACATGCGTGGATTTTGTAGCTTGGTTACA 261

Qy   1022 CAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACACT 1081
            ||  | | | |          |||||   || |   || ||     |    ||   | |
Db    262 GGAATCTCTTAAAAAGATGATTGCAGATTTGGCTGCTTAGGATATACTTTTGCAGTTATT 321

Qy   1082 GCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATCC 1141
            || | ||||||||  |   |||||  ||   |  |||       |  | | |||  |
Db    322 AGCTTTTGACAATGCTAATCACTATTGCTAGGCTGCTCTGCGACCTATCAGAGGGAAAGC 381

Qy   1142 CACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCTAG 1201
            |||  ||     | | ||| ||| |  |  |  | ||      ||| || ||||
Db    382 ACATTTAGTTGATTATATCAAGGCCTGTGATAGTATCAGAGGTAATCTACATAAAGCTAC 441

Qy   1202 ATTAATGGCAGAAGCCCTGAAAG 1224
            ||  ||||| | ||  ||  ||
Db    442 TTTGTTGGCACAGGCGATGGCAG 464


RESULT 12
AW629027/c
LOCUS       AW629027                571 bp    mRNA    linear   EST 31-MAR-2000
DEFINITION  hi46g08.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:2975390 3' similar to SW:GAG2_HUMAN P10264 RETROVIRUS-RELATED
            GAG POLYPROTEIN ;, mRNA sequence.
ACCESSION   AW629027
VERSION     AW629027.1  GI:7375817
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 571)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40UP from Gibco
            High quality sequence stop: 444.
FEATURES             Location/Qualifiers
     source          1. .571
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2975390"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo. "
BASE COUNT      142 a    151 c    109 g    168 t      1 others
ORIGIN

  Query Match             2.7%;  Score 51.8;  DB 10;  Length 571;
  Best Local Similarity   49.8%;  Pred. No. 0.19;
  Matches  131;  Conservative    0;  Mismatches  132;  Indels    0;  Gaps    0;

Qy    962 CATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCTA 1021
          | ||   |  |||||||| || |||||||| || |         | | |||    |||  |
Db    382 CTTTAGTGCAGTAAAACAGGGACCAAAAGAACCGTACGCAGATTTTATAGCTCAGTTACA 323

Qy   1022 CAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACACT 1081
            |   | | | |   |       ||||| | || |   || ||    ||   |||  |
Db    322 GGAGTCTCTTAAAAAGGTGATTGCAGATTCGGCTGCTCAGGATATAGTGCTGCAGTTATC 263

Qy   1082 GCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATCC 1141
            |  | |||||||| || |||||| ||   |  |||       |  | | |||  |
Db    262 AGCTTTCGACAATGCTAATCCCGATTGCCAGGCTGCTCTGTGACCTATCAGAGGGAACGC 203

Qy   1142 CACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCTAG 1201
            | |||  ||     | | ||| ||| | ||  |  | ||       || || ||||
Db    202 ACACTTAGTTGATTATATCAAGGCCTGTGACGGTATCAGAGGTAATATGCATAAAGCTAC 143

Qy   1202 ATTAATGGCAGAAGCCCTGAAAG 1224
            ||  ||| | | ||  ||  ||
Db    142 TTTGTTGGTACAGGCAATGGCAG 120


RESULT 13
BM468018
LOCUS       BM468018                1300 bp    mRNA    linear   EST 05-FEB-2002
DEFINITION  AGENCOURT_6437618 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5532913
```

```
            5', mRNA sequence.
ACCESSION   BM468018
VERSION     BM468018.1  GI:18517060
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1300)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12216  row: p  column: 02
            High quality sequence stop: 434.
FEATURES             Location/Qualifiers
     source          1. .1300
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5532913"
                     /clone_lib="NIH_MGC_71"
                     /tissue_type="leiomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 2.1 kb. "
BASE COUNT      586 a    102 c    112 g    112 t    388 others
ORIGIN
```

```
  Query Match             2.6%;  Score 50.6;  DB 13;  Length 1300;
  Best Local Similarity   27.4%;  Pred. No. 0.46;
  Matches  164;  Conservative    0;  Mismatches  434;  Indels    0;  Gaps    0;

Qy      26 AGAGGAAGAGGCCTCCGGTTGCAGGTAAGTGCAACACAAAAAAGAAATAGCTGTCTTTTA 85
           |   || |           |          || | |||||| | | |          |
Db     619 ANNNNAAAAAAANNANNNNNNAANNNNNANNAAANAAAAAAAAAAANAAAANNNNNNNNA 678

Qy      86 TCCAGGAAGGGGTAATAAGATAGAGTGGGAGATGGGCGTGAGAAACTCCGTCTTGTCAGG 145
           |  ||     || ||          | |          | |||          |
Db     679 AAAAAAAANNAAAAAAAANNNNNNNNNNNAAAAAAANNAAAAAAANNNNANAAAANNAAN 738

Qy     146 GAAGAAAGCAGATGAATTAGAAAAAATTAGGCTACGACCCAACGGAAAGAAAAAGTACAT 205
           ||        |  ||  | |||| |  |          |    || |||||
Db     739 NAANNNNNNNNANNAANNAAAAAANANAANAANNAANAAANANNNNAAAAAAAANNNNNN 798

Qy     206 GTTGAAGCATGTAGTATGGGCAGCAAATGAATTAGATAGATTTGGATTAGCAGAAAGCCT 265
           |  |       |     |  |||   |  |    |       |  |    |||
Db     799 NNANNAAAAAAANNAAAAAAAAAAAAAANNNAAAANNNAANAAAAANAAANNNAAAANNN 858

Qy     266 GTTGGAGAACAAAGAAGGATGTCAAAAAATACTTTCGGTCTTAGCTCCATTAGTGCCAAC 325
           |  ||  ||     |   |||||  |                          |    ||
Db     859 NAAANAAAAANAANNNNAANNNNNAAAAAAANNNNNAAAANNNNNNNNNNNANNAAAAAA 918

Qy     326 AGGCTCAGAAAATTTAAAAAGCCTTTATAATACTGTCTGCGTCATCTGGTGCATTCACGC 385
           |       |||     ||||       || |          |      |   |
Db     919 AAAANNNAAAANANNNAAAANNNNANNNAAAANNNANANAAAANAAAANNNNNAANAANNA 978

Qy     386 AGAAGAGAAAGTGAAACACACTGAGGAAGCAAAACAGATAGTGCAGAGACACCTAGTGGT 445
           | || | |||           ||  |||    | |    |          |
Db     979 AAAAAAAAAAANNNNNANANNNNNNAAAAAAAANNNNAAANNNNNNAANANAAAAAANA 1038

Qy     446 GGAAACAGGAACAACAGAAACTATGCCAAAAACAAGTAGACCAACAGCACCATCTAGCGG 505
           | |  || |  | ||   |    ||| |    |    || |  |      |
Db    1039 NANNAAAAAAANANAAAAANAAAAANAAAANANNNNAANNAAAAAANAANNNANNAAAAA 1098
```

```
Qy     506 CAGAGGAGGAAATTACCCAGTACAACAAATAGGTGGTAACTATGTCCACCTGCCATTAAG 565
           | |  |  |||  |    |  | ||   | |        || |      |    |  |
Db    1099 AAAAAAAAAAAAAAAANAAAAAAAAAAANNANAAAAAANAANNANAAAAAAAAAAANNANA 1158

Qy     566 CCCGAGAACATTAAATGCCTGGGTAAAATTGATAGAGGAAAAGAAATTTGGAGCAGAA 623
           |  ||    |          |  |  | | |  | |   ||          |
Db    1159 NAAAAAAAANAAANNNANANNNANANNAAAAAAAAAANNANANNNAANNANNNNNNNA 1216
```

```
RESULT 14
R93353/c
LOCUS       R93353                   496 bp    mRNA    linear   EST 22-JAN-1999
DEFINITION  yt69a11.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
            IMAGE:275612 5' similar to SP:GAG2_HUMAN P10264 RETROVIRUS-RELATED
            GAG POLYPROTEIN ;, mRNA sequence.
ACCESSION   R93353
VERSION     R93353.1  GI:967519
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 496)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 1133
            High quality sequence stops: 214
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Possible reversed clone: similarity on wrong strand
            Insert Length: 1133    Std Error: 0.00
            Seq primer: M13RP1
            High quality sequence stop: 214.
FEATURES             Location/Qualifiers
     source          1. .496
                     /organism="Homo sapiens"
                     /db_xref="GDB:3799130"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:275612"
                     /clone_lib="Soares fetal liver spleen 1NFLS"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Pac I - oligo(dT) primer
                     [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      129 a    117 c     97 g    149 t      4 others
ORIGIN
```

```
  Query Match             2.6%;  Score 50.2;  DB 14;  Length 496;
  Best Local Similarity   49.4%;  Pred. No. 0.42;
  Matches  130;  Conservative    0;  Mismatches  133;  Indels    0;  Gaps    0;

Qy     962 CATTCTAGATGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCTA 1021
           | ||   |  |||||||| || |||||| | ||||    |   | | |||   |||| |
Db     327 CTTTAGTGCAGTAAAACAGGGACCAAAAAAACCATACATGGATTTTATAGCTCGGTTACA 268
```

```
Qy   1022 CAAAAGTTTAAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACACT 1081
          |   | | | |   | |    ||||| |  | |   ||||    ||   ||   | |
Db    267 GGAGTCTCTTAAAAAGGTAACTGCAGATTCGACTGCTCAGGATATAGTGTTGCACTTATT 208

Qy   1082 GCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATCC 1141
            |  | |||||||| || || |||  |  ||  |||       |  | | |||  |
Db    207 AGCTTTCAACAATGCTAATCCCGACTGCCTGGCAGCTCTGCGACCTATCAGAGGGAAAGC 148

Qy   1142 CACCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCTAG 1201
            |||  ||       | ||| ||| | ||  | ||          || |||||||
Db    147 ACATTTAGTTGATTATACCAAGGCCTGTGATGGTATTGGAATTAATCTGCATAAGGCTAC 88

Qy   1202 ATTAATGGCAGAAGCCCTGAAAG 1224
            |  | ||| | ||  ||  ||
Db     87 TCTGCTAGCACAGGCAATGGCAG 65


RESULT 15
B56450
LOCUS       B56450                   491 bp    DNA     linear   GSS 20-JUN-1998
DEFINITION  CIT-HSP-2006O23.TR CIT-HSP Homo sapiens genomic clone 2006O23, DNA
            sequence.
ACCESSION   B56450
VERSION     B56450.1  GI:2610784
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 491)
  AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden
            ,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M.
            and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: CIT-HSP-2006O23.TV
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .491
                     /organism="Homo sapiens"
                     /db_xref="GDB:7039849"
                     /db_xref="taxon:9606"
                     /clone="2006O23"
                     /clone_lib="CIT-HSP"
                     /sex="Male"
                     /cell_type="Sperm"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
BASE COUNT      135 a     91 c    129 g    136 t
ORIGIN

  Query Match             2.6%;  Score 49.2;  DB 17;  Length 491;
  Best Local Similarity   49.6%;  Pred. No. 0.72;
  Matches  126;  Conservative    0;  Mismatches  128;  Indels    0;  Gaps    0;

Qy    971 TGTAAAACAAGGGCCAAAAGAGCCATTTCAGAGCTATGTAGACAGGTTCTACAAAAGTTT 1030
          ||||||||| || |||| ||| ||||    |   | | |||   ||||  |  |   | |
Db    186 TGTAAAACAGGGACCAAGAGAACCATACACGGATTTTATAGCTTGGTTACAGGAGTCTCT 245

Qy   1031 AAGAGCAGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACACTGCTGATTCA 1090
            | |   |      ||||| | || |   || ||   ||        | |     | |
Db    246 TAAAAAGGTGATTGCAGATTCGGCTGCTCAGGATATAGTGTTGTGGTTATTAGCTTTCGA 305

Qy   1091 AAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAATCCCACCCTAGA 1150
            |||||||| || |||||| ||   |  |||       |  | | |||  |     |||
Db    306 CAATGCTAATCCCGATTGCCAGGCTGCTCTGCGACCTATCAGAGGGAAAGCACATTTAGT 365

Qy   1151 AGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGGCCGGGACAGAAGGCTAGATTAATGGC 1210
            ||     | | ||| ||| | ||  | || |        || |||||||   |  | ||
Db    366 TGATTATATCAAGGCCTGTGATGGTATTGGAGATAATCTGCATAAGGCTACTCTGCTAGC 425

Qy   1211 AGAAGCCCTGAAAG 1224
            | |   ||  ||
Db    426 CCAGGTAATGGCAG 439


Search completed: November 16, 2002, 13:06:45
Job time : 1798.71 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 16, 2002, 07:31:18 ; Search time 4381.28 Seconds
(without alignments)
17031.470 Million cell updates/sec

Title: M33262_COPY_3105_5668
Perfect score: 2564
Sequence: 1 TTTCCCATAGCTAAAGTAGA.....GCTAGAGAGGTGGCATAGCC 2564

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2564	100.0	9860	14	SIU89134	U89134 Simian-Huma
2	2564	100.0	10000	14	AF038398	AF038398 Simian-Hu
3	2564	100.0	10006	14	AF038399	AF038399 Simian-Hu
4	2564	100.0	10006	14	AF041850	AF041850 Simian-Hu
5	2564	100.0	10279	14	AY033146	AY033146 Simian im
6	2564	100.0	10279	14	AY033233	AY033233 Simian im
7	2564	100.0	10535	14	SIVMM239	M33262 Simian (mac
8	2559.2	99.8	9860	14	AF217181	AF217181 Simian-Hu
9	2519.8	98.3	10277	14	SIVMM32H	D01065 Simian immu
10	2515	98.1	10277	14	SIVMM251	M19499 Simian (mac
11	2510.2	97.9	10274	14	SIV1A11AA	M76764 Simian immu
12	2470.2	96.3	9646	14	RESIVMXX	Y00277 Simian immu
13	2455.2	95.8	9664	14	SIU79412	U79412 Simian immu
14	2453.6	95.7	9628	14	SIVMNE	M32741 Simian immu
15	2004	78.2	10289	14	SIU72748	U72748 Simian immu
16	2003	78.1	10184	14	AF077017	AF077017 Simian im
17	2001.4	78.1	9996	14	SIVSMMPBJA	M31325 Simian immu
18	1993.6	77.8	10241	14	RESIVSMM	X14307 Simian immu
19	1993.4	77.7	9996	14	SIVSMMPBJB	L03295 Simian immu
20	1993.4	77.7	10246	14	SIV6P12	L09211 Simian immu
21	1993.4	77.7	10246	14	SIV6P6	L09212 Simian immu
22	1993.4	77.7	10246	14	SIV6P9	L09213 Simian immu
23	1973.8	77.0	9675	14	SIVGAGAA	M80193 Simian immu
24	1963.8	76.6	9638	14	SIVGAGAB	M80194 Simian immu
25	1904.8	74.3	9892	14	SIVSTM	M83293 Simian (stu
26	1759.8	68.6	9525	14	HIV2MDS	Z48731 Human immun
27	1753.4	68.4	10163	6	AR153663	AR153663 Sequence
28	1753.4	68.4	10172	14	HIV2U22047	U22047 Human immun
29	1751.4	68.3	3168	6	AR153668	AR153668 Sequence
30	1747	68.1	11443	14	HIV2CAM2	D00835 Human immun
31	1742	67.9	9599	14	AF208027	AF208027 HIV-2 iso
32	1738.4	67.8	9636	6	A05350	A05350 HIV-2 (SBL/
33	1738.4	67.8	9636	14	HIV2ISY	J04498 Human immun
34	1734.2	67.6	9671	14	HIV2ROD	M15390 Human immun
35	1734.2	67.6	9671	14	HIV2RODX	X05291 Human immun
36	1733.6	67.6	10312	14	HIV2U38293	U38293 Human immun
37	1732	67.6	10359	14	HIV2BEN	M30502 Human immun
38	1723	67.2	9672	6	I09513	I09513 Sequence 1
39	1723	67.2	9672	14	HIV2ST	M31113 Human immun
40	1717.6	67.0	9477	6	E02138	E02138 cDNA to gRN
41	1715	66.9	9472	14	HIV2D194	J04542 Human immun
42	1715	66.9	9472	14	HIVD194	X52223 Human immun
43	1714.4	66.9	9480	14	HIV2GH1	M30895 Human immun
44	1713.4	66.8	9473	6	A09995	A09995 HIV-2 (HIV-
45	1708.6	66.6	10353	14	AF082339	AF082339 HIV-2 iso

ALIGNMENTS

RESULT 1
SIU89134
LOCUS SIU89134 9860 bp DNA linear VRL 21-MAY-1997
DEFINITION Simian-Human immunodeficiency virus strain SHIV-89.6P, complete genome.
ACCESSION U89134
VERSION U89134.1 GI:2108163
KEYWORDS .
SOURCE Simian-Human immunodeficiency virus.
ORGANISM Simian-Human immunodeficiency virus
Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate lentivirus group.
REFERENCE 1 (bases 1 to 9860)
AUTHORS Karlsson,G.B., Halloran,M., Li,J., Park,I.W., Gomila,R., Reimann,K.A., Axthelm,M.K., Iliff,S.A., Letvin,N.L. and Sodroski,J.

```
  TITLE     Characterization of molecularly cloned simian-human
            immunodeficiency viruses causing rapid CD4+ lymphocyte depletion in
            rhesus monkeys
  JOURNAL   J. Virol. 71 (6), 4218-4225 (1997)
  MEDLINE   97296224
   PUBMED   9151808
REFERENCE   2  (bases 1 to 9860)
  AUTHORS   Karlsson,G.B., Halloran,M., Li,J., Park,I., Gomila,R.,
            Reimann,K.A., Axthelm,M.K., Iliff,S.A., Letvin,N.L. and Sodroski,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-FEB-1997) Human Retrovirology, Dana-Farber Cancer
            Institute, 44 Binney Street, Boston, MA 02115, USA
FEATURES             Location/Qualifiers
     source          1. .9860
                     /organism="Simian-Human immunodeficiency virus"
                     /proviral
                     /strain="SHIV-89.6P"
                     /db_xref="taxon:57667"
                     /clone="SHIV-89.6KB9"
     source          1. .5946
                     /organism="Simian-Human immunodeficiency virus"
                     /db_xref="taxon:57667"
                     /note="derived from SIV"
     source          5947. .8778
                     /organism="Simian-Human immunodeficiency virus"
                     /db_xref="taxon:57667"
                     /note="derived from HIV-1"
     source          8779. .9860
                     /organism="Simian-Human immunodeficiency virus"
                     /db_xref="taxon:57667"
                     /note="derived from SIV"
     LTR             1. .301
                     /note="5' LTR"
     misc_feature    1. .177
                     /note="R region"
     misc_feature    178. .301
                     /note="U5 region"
     gene            536. .2068
                     /gene="gag"
     CDS             536. .2068
                     /gene="gag"
                     /codon_start=1
                     /product="gag polyprotein"
                     /protein_id="AAC57419.1"
                     /db_xref="GI:2108164"
                     /translation="MGVRNSVLSGKKADELEKIRLRPNGKKKYMLKHVVWAANELDRF
                     GLAESLLENKEGCQKILSVLAPLVPTGSENLKSLYNTVCVIWCIHAEEKVKHTEEAKQ
                     IVQRHLVVETGTTETMPKTSRPTAPSSGRGGNYPVQQIGGNYVHLPLSPRTLNAWVKL
                     IEEKKFGAEVVPGFQALSEGCTPYDINQMLNCVGDHQAAMQIIRDIINEEAADWDLQH
                     PQPAPQQGQLREPSGSDIAGTTSSVDEQIQWMYRQQNPIPVGNIYRRWIQLGLQKCVR
                     MYNPTNILDVKQGPKEPFQSYVDRFYKSLRAEQTDAAVKNWMTQTLLIQNANPDCKLV
                     LKGLGVNPTLEEMLTACQGVGGPGQKARLMAEALKEALAPVPIPFAAAQQRGPRKPIK
                     CWNCGKEGHSARQCRAPRRQGCWKCGKMDHVMAKCPDRQAGFLGLGPWGKKPRNFPMA
                     QVHQGLMPTAPPEDPAVDLLKNYMQLGKQQREKQRESREKPYKEVTEDLLHLNSLFGG
                     DQ"
     gene            <1711. .4893
                     /gene="pol"
     CDS             <1714. .4893
                     /gene="pol"
                     /codon_start=1
                     /product="pol polyprotein"
                     /protein_id="AAC57420.1"
                     /db_xref="GI:2108165"
                     /translation="LELWERGTLCKAMQSPKKTGMLEMWKNGPCYGQMPRQTGGFFRP
                     WSMGKEAPQFPHGSSASGADANCSPRGPSCGSAKELHAVGQAAERKAERKQREALQGG
                     DRGFAAPQFSLWRRPVVTAHIEGQPVEVLLDTGADDSIVTGIELGPHYTPKIVGGIGG
                     FINTKEYKNVEIEVLGKRIKGTIMTGDTPINIFGRNLLTALGMSLNFPIAKVEPVKVA
                     LKPGKDGPKLKQWPLSKEKIVALREICEKMEKDGQLEEAPPTNPYNTPTFAIKKKDKN
                     KWRMLIDFRELNRVTQDFTEVQLGIPHPAGLAKRKRITVLDIGDAYFSIPLDEEFRQY
                     TAFTLPSVNNAEPGKRYIYKVLPQGWKGSPAIFQYTMRHVLEPFRKANPDVTLVQYMD
                     DILIASDRTDLEHDRVVLQSKELLNSIGFSTPEEKFQKDPPFQWMGYELWPTKWKLQK
                     IELPQRETWTVNDIQKLVGVLNWAAQIYPGIKTKHLCRLIRGKMTLTEEVQWTEMAEA
                     EYEENKIILSQEQEGCYYQEGKPLEATVIKSQDNQWSYKIHQEDKILKVGKFAKIKNT
                     HTNGVRLLAHVIQKIGKEAIVIWGQVPKFHLPVEKDVWEQWWTDYWQVTWIPEWDFIS
                     TPPLVRLVFNLVKDPIEGEETYYTDGSCNKQSKEGKAGYITDRGKDKVKVLEQTTNQQ
                     AELEAFLMALTDSGPKANIIVDSQYVMGIITGCPTESESRLVNQIIEEMIKKSEIYVA
                     WVPAHKGIGGNQEIDHLVSQGIRQVLFLEKIEPAQEEHDKYHSNVKELVFKFGLPRIV
                     ARQIVDTCDKCHQKGEAIHGQANSDLGTWQMDCTHLEGKIIIVAVHVASGFIEAEVIP
                     QETGRQTALFLLKLAGRWPITHLHTDNGANFASQEVKMVAWWAGIEHTFGVPYNPQSQ
                     GVVEAMNHHLKNQIDRIREQANSVETIVLMAVHCMNFKRRGGIGDMTPAERLINMITT
                     EQEIQFQQSKNSKFKNFRVYYREGRDQLWKGPGELLWKGEGAVILKVGTDIKVVPRRK
                     AKIIKDYGGGKEVDSSSHMEDTGEAREVA"
     gene            4823. .5467
                     /gene="vif"
     CDS             4823. .5467
                     /gene="vif"
                     /codon_start=1
                     /product="vif protein"
                     /protein_id="AAC57421.1"
                     /db_xref="GI:2108166"
                     /translation="MEEEKRWIAVPTWRIPERLERWHSLIKYLKYKTKDLQKVCYVPH
                     FKVGWAWWTCSRVIFPLQEGSHLEVQGYWHLTPEKGWLSTYAVRITWYSKNFWTDVTP
                     NYADILLHSTYFPCFTAGEVRRAIRGEQLLSCCRFPRAHKYQVPSLQYLALKVVSDVR
                     SQGENPTWKQWRRDNRRGLRMAKQNSRGDKQRGGKPPTKGANFPGLAKVLGILA"
     gene            5295. .5633
                     /gene="vpx"
     CDS             5295. .5633
                     /gene="vpx"
                     /codon_start=1
                     /product="vpx protein"
                     /protein_id="AAC57422.1"
                     /db_xref="GI:2108167"
                     /translation="MSDPRERIPPGNSGEETIGEAFEWLNRTVEEINREAVNHLPREL
                     IFQVWQRSWEYWHDEQGMSPSYVKYRYLCLIQKALFMHCKKGCRCLGEGHGAGGWRPG
                     PPPPPPPGLA"
     gene            5634. .5939
                     /gene="vpr"
     CDS             5634. .5939
                     /gene="vpr"
                     /codon_start=1
                     /product="vpr protein"
                     /protein_id="AAC57423.1"
                     /db_xref="GI:2108168"
                     /translation="MEERPPENEGPQREPWDEWVVEVLEELKEEALKHFDPRLLTALG
                     NHIYNRHGDTLEGAGELIRILQRALFMHFRGGCIHSRIGQPGGGNPLSAIPPSRSML"
     gene            5948. .8580
                     /gene="tat"
     CDS             join(5948. .6162,8487. .8580)
                     /gene="tat"
                     /codon_start=1
                     /product="tat protein"
                     /protein_id="AAC57424.1"
                     /db_xref="GI:2108169"
                     /translation="MEPVDPRLEPWKHPGSKPKTACTNCYCKKCCFHCQVCFTTKALG
                     ISYGRKKRRQRRRAHQNSQTHQASLSKQPSSQPRGDPTGPKEQKKKVERETETDPVHQ
                     "
     gene            6087. .8761
                     /gene="rev"
     CDS             join(6087. .6162,8487. .8761)
                     /gene="rev"
                     /codon_start=1
                     /product="rev protein"
                     /protein_id="AAC57425.1"
                     /db_xref="GI:2108170"
                     /translation="MAGRSGDSDEELIRTVRLIKLLYQSNPPPSLEGTRQARRNRRRR
                     WRERQRQIRSISERILGTYLGRSAEPVPLQLPPLERLTLDCNEDCGTSGTQGVGSPQI
                     LVESPTVLESGTKE"
     gene            6179. .6424
                     /gene="vpu"
     CDS             6179. .6424
                     /gene="vpu"
                     /codon_start=1
                     /product="vpu protein"
                     /protein_id="AAC57426.1"
                     /db_xref="GI:2108171"
                     /translation="MQPIQIAIVALVVAIIIAIVVWSIVIIEYRKILRQRKIDRLIDR
```

```
                     LIERAEDSGNESEGEISALVEMGVEMGHHAPWDVDDL"
     gene            6339. .8948
                     /gene="env"
     CDS             6339. .8948
                     /gene="env"
                     /codon_start=1
                     /product="env polyprotein"
                     /protein_id="AAC57427.1"
                     /db_xref="GI:2108172"
                     /translation="MRVKEKYQHLWRWGWRWGTMLLGMLMICSATEKLWVTVYYGVPV
                     WREATTTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEVVLGNVTENFNMWKNNMVD
                     QMHEDIISLWDESLKPCVKLTPLCVTLNCTNLNITKNTTNLTSSSWGMMEEGEIKNCS
                     FYITTSIRNKVKKEYALFNRLDVVPVKNTSNTKYRLISCNTSVITQACPKVSFQPIPI
                     HYCVPAGFAILKCNNKTFNGSGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAEEDIVIR
                     SEDFTDNVKTIIVQLNESVVINCTRPNNNTRERLSIGPGRAFYARRNIIGDIRQAHCN
                     ISRAKWNNTLQQIVIKLREKFRNKTIAFNQSSGGDPEIVMHSFNCGGEFFYCNTAQLF
                     NSTWNVAGGTNGTEGNDIITLQCRIKQIINMWQKVGKAMYAPPITGQIRCSSNITGLL
                     LTRDGGNSTETETEIFRPGGGDMRDNWRSELYKYKVVRIEPIGVAPTRAKRRTVQREK
                     RAVGIGAVFLGFLGAAGSTMGAASVTLTVQARLLLSGIVQQQNNLLRAIEAQQNMLRL
                     TVWGIKQLQARVLALERYLRDQQLMGIWGCSGKLICTTSVPWNVSWSNKSVDDIWNNM
                     TWMEWEREIDNYTDYIYDLLEKSQTQQEKNEKELLELDKWASLWNWFDITNWLWYIRL
                     FIMIVGGLIGLRIVFAVLSIVNRVRQGYSPLSFQTLLPASRGPDRPEGTEEEGGERDR
                     DRSGPSVNGSLALIWDDLRSLCLFSYHRLRDLLLIVTRIVELLGRRGWEALKYWWNLL
                     QYWSQELKNSAVSLLQYGWSYFHEAVQAVWRSATETLAGAWGDLWETLRRGGRWILAI
                     PRRIRQGLELTLL"
     gene            8782. .9573
                     /gene="nef"
     CDS             8782. .9573
                     /gene="nef"
                     /codon_start=1
                     /product="nef protein"
                     /protein_id="AAC57428.1"
                     /db_xref="GI:2108173"
                     /translation="MGGAISMRRSRPSGDLRQRLLRARGETYGRLLGEVEDGYSQSPG
                     GLDKGLSSLSCEGQKYNQGQYMNTPWRNPAEEREKLAYRKQNMDDIDEEDDDLVGVSV
                     RPKVPLRTMSYKLAIDMSHFIKEKGGLEGIYYSARRHRILDIYLEKEEGIIPDWQDYT
                     SGPGIRYPKTFGWLWKLVPVNVSDEAQEDEEHYLMHPAQTSQWDDPWGEVLAWKFDPT
                     LAYTYEAYVRYPEEFGSKSGLSEEEVRRRLTARGLLNMADKKETR"
     LTR             9167. .9860
                     /note="3' LTR"
     misc_feature    9167. .9683

  Query Match           100.0%;  Score 2564;  DB 14;  Length 9860;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 2564;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2332 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 2391

Qy      61 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2392 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 2451

Qy     121 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2452 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 2511

Qy     181 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2512 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 2571

Qy     241 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2572 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 2631

Qy     301 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2632 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 2691

Qy     361 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2692 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 2751

Qy     421 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2752 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 2811

Qy     481 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2812 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 2871

Qy     541 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2872 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 2931

Qy     601 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2932 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 2991

Qy     661 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2992 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 3051

Qy     721 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3052 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 3111

Qy     781 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3112 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 3171

Qy     841 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3172 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 3231

Qy     901 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3232 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 3291

Qy     961 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3292 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 3351

Qy    1021 AAAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3352 AAAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACA 3411

Qy    1081 CATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATA 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3412 CATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATA 3471

Qy    1141 GTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3472 GTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAG 3531

Qy    1201 TGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCA 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3532 TGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCA 3591

Qy    1261 CCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTAT 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3592 CCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTAT 3651

Qy    1321 TATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGAT 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3652 TATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGAT 3711

Qy    1381 AGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAA 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3712 AGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAA 3771

Qy    1441 GCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAA 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3772 GCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAA 3831
```

```
Qy   1501 TATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAA 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3832 TATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAA 3891

Qy   1561 ATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAA 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3892 ATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAA 3951

Qy   1621 GGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTC 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3952 GGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTC 4011

Qy   1681 TTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4012 TTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAA 4071

Qy   1741 GAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGT 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4072 GAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGT 4131

Qy   1801 GATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACT 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4132 GATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACT 4191

Qy   1861 TGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCT 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4192 TGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCT 4251

Qy   1921 AGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTT 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4252 AGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTT 4311

Qy   1981 CTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAAC 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4312 CTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAAC 4371

Qy   2041 TTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGG 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4372 TTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGG 4431

Qy   2101 GTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAAT 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4432 GTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAAT 4491

Qy   2161 CAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTT 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4492 CAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTT 4551

Qy   2221 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTA 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4552 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTA 4611

Qy   2281 ATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTT 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4612 ATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTT 4671

Qy   2341 AAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAG 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4672 AAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAG 4731

Qy   2401 CTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTA 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4732 CTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTA 4791

Qy   2461 CCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGT 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4792 CCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGT 4851

Qy   2521 TCCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 2564
          |||||||||||||||||||||||||||||||||||||||||||
Db   4852 TCCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 4895
```

```
RESULT 2
AF038398
LOCUS       AF038398               10000 bp    DNA     linear   VRL 02-FEB-1998
DEFINITION  Simian-Human immunodeficiency virus strain SHIV-89.6, complete
            genome.
ACCESSION   AF038398
VERSION     AF038398.1  GI:2828036
KEYWORDS    .
SOURCE      Simian-Human immunodeficiency virus.
  ORGANISM  Simian-Human immunodeficiency virus
            Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
            lentivirus group.
REFERENCE   1  (bases 1 to 10000)
  AUTHORS   Reimann,K.A., Li,J.T., Voss,G., Lekutis,C., Tenner-Racz,K.,
            Racz,P., Lin,W., Montefiori,D.C., Lee-Parritz,D.E., Lu,Y.,
            Collman,R.G., Sodroski,J. and Letvin,N.L.
  TITLE     An env gene derived from a primary human immunodeficiency virus
            type 1 isolate confers high in vivo replicative capacity to a
            chimeric simian/human immunodeficiency virus in rhesus monkeys
  JOURNAL   J. Virol. 70 (5), 3198-3206 (1996)
  MEDLINE   96186751
   PUBMED   8627800
REFERENCE   2  (bases 1 to 10000)
  AUTHORS   Reimann,K.A., Li,J.T., Voss,G., Lekutis,C., Tenner-Racz,K.,
            Racz,P., Lin,W., Montefiori,D.C., Lee-Parritz,D.E., Lu,Y.,
            Collman,R.G., Sodroski,J. and Letvin,N.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-DEC-1997) Human Retrovirology, Dana-Farber Cancer
            Institute, 44 Binney Street, Boston, MA 02115, USA
FEATURES             Location/Qualifiers
     source          1. .10000
                     /organism="Simian-Human immunodeficiency virus"
                     /proviral
                     /strain="SHIV-89.6"
                     /db_xref="taxon:57667"
     LTR             1. .301
                     /note="5' long terminal repeat"
     repeat_region   1. .177
                     /note="R region"
     misc_feature    178. .301
                     /note="U5 region"
     gene            536. .2068
                     /gene="gag"
     CDS             536. .2068
                     /gene="gag"
                     /codon_start=1
                     /product="gag"
                     /protein_id="AAB99958.1"
                     /db_xref="GI:2828037"
                     /translation="MGVRNSVLSGKKADELEKIRLRPNGKKKYMLKHVVWAANELDRF
                     GLAESLLENKEGCQKILSVLAPLVPTGSENLKSLYNTVCVIWCIHAEEKVKHTEEAKQ
                     IVQRHLVVETGTTETMPKTSRPTAPSSGRGGNYPVQQIGGNYVHLPLSPRTLNAWVKL
                     IEEKKFGAEVVPGFQALSEGCTPYDINQMLNCVGDHQAAMQIIRDIINEEAADWDLQH
                     PQPAPQQGQLREPSGSDIAGTTSSVDEQIQWMYRQQNPIPVGNIYRRWIQLGLQKCVR
                     MYNPTNILDVKQGPKEPFQSYVDRFYKSLRAEQTDAAVKNWMTQTLLIQNANPDCKLV
                     LKGLGVNPTLEEMLTACQGVGGPGQKARLMAEALKEALAPVPIPFAAAQQRGPRKPIK
                     CWNCGKEGHSARQCRAPRRQGCWKCGKMDHVMAKCPDRQAGFLGLGPWGKKPRNFPMA
                     QVHQGLMPTAPPEDPAVDLLKNYMQLGKQQREKQRESREKPYKEVTEDLLHLNSLFGG
                     DQ"
     gene            <1714. .4893
                     /gene="pol"
     CDS             <1714. .4893
                     /gene="pol"
                     /codon_start=1
                     /product="pol"
                     /protein_id="AAB99959.1"
                     /db_xref="GI:2828038"
                     /translation="LELWERGTLCKAMQSPKKTGMLEMWKNGPCYGQMPRQTGGFFRP
                     WSMGKEAPQFPHGSSASGADANCSPRGPSCGSAKELHAVGQAAERKAERKQREALQGG
                     DRGFAAPQFSLWRRPVVTAHIEGQPVEVLLDTGADDSIVTGIELGPHYTPKIVGGIGG
                     FINTKEYKNVEIEVLGKRIKGTIMTGDTPINIFGRNLLTALGMSLNFPIAKVEPVKVA
                     LKPGKDGPKLKQWPLSKEKIVALREICEKMEKDGQLEEAPPTNPYNTPTFAIKKKDKN
                     KWRMLIDFRELNRVTQDFTEVQLGIPHPAGLAKRKRITVLDIGDAYFSIPLDEEFRQY
```

```
                TAFTLPSVNNAEPGKRYIYKVLPQGWKGSPAIFQYTMRHVLEPFRKANPDVTLVQYMD
                DILIASDRTDLEHDRVVLQSKELLNSIGFSTPEEKFQKDPPFQWMGYELWPTKWKLQK
                IELPQRETWTVNDIQKLVGVLNWAAQIYPGIKTKHLCRLIRGKMTLTEEVQWTEMAEA
                EYEENKIILSQEQEGCYYQEGKPLEATVIKSQDNQWSYKIHQEDKILKVGKFAKIKNT
                HTNGVRLLAHVIQKIGKEAIVIWGQVPKFHLPVEKDVWEQWWTDYWQVTWIPEWDFIS
                TPPLVRLVFNLVKDPIEGEETYYTDGSCNKQSKEGKAGYITDRGKDKVKVLEQTTNQQ
                AELEAFLMALTDSGPKANIIVDSQYVMGIITGCPTESESRLVNQIIEEMIKKSEIYVA
                WVPAHKGIGGNQEIDHLVSQGIRQVLFLEKIEPAQEEHDKYHSNVKELVFKFGLPRIV
                ARQIVDTCDKCHQKGEAIHGQANSDLGTWQMDCTHLEGKIIIVAVHVASGFIEAEVIP
                QETGRQTALFLLKLAGRWPITHLHTDNGANFASQEVKMVAWWAGIEHTFGVPYNPQSQ
                GVVEAMNHHLKNQIDRIREQANSVETIVLMAVHCMNFKRRGGIGDMTPAERLINMITT
                EQEIQFQQSKNSKFKNFRVYYREGRDQLWKGPGELLWKGEGAVILKVGTDIKVVPRRK
                AKIIKDYGGGKEVDSSSHMEDTGEAREVA"
gene            4823. .5467
                /gene="vif"
CDS             4823. .5467
                /gene="vif"
                /codon_start=1
                /product="vif"
                /protein_id="AAB99960.1"
                /db_xref="GI:2828039"
                /translation="MEEEKRWIAVPTWRIPERLERWHSLIKYLKYKTKDLQKVCYVPH
                FKVGWAWWTCSRVIFPLQEGSHLEVQGYWHLTPEKGWLSTYAVRITWYSKNFWTDVTP
                NYADILLHSTYFPCFTAGEVRRAIRGEQLLSCCRFPRAHKYQVPSLQYLALKVVSDVR
                SQGENPTWKQWRRDNRRGLRMAKQNSRGDKQRGGKPPTKGANFPGLAKVLGILA"
gene            5295. .5633
                /gene="vpx"
CDS             5295. .5633
                /gene="vpx"
                /codon_start=1
                /product="vpx"
                /protein_id="AAB99961.1"
                /db_xref="GI:2828040"
                /translation="MSDPRERIPPGNSGEETIGEAFEWLNRTVEEINREAVNHLPREL
                IFQVWQRSWEYWHDEQGMSPSYVKYRYLCLIQKALFMHCKKGCRCLGEGHGAGGWRPG
                PPPPPPPGLA"
gene            5634. .5939
                /gene="vpr"
CDS             5634. .5939
                /gene="vpr"
                /codon_start=1
                /product="vpr"
                /protein_id="AAB99962.1"
                /db_xref="GI:2828041"
                /translation="MEERPPENEGPQREPWDEWVVEVLEELKEEALKHFDPRLLTALG
                NHIYNRHGDTLEGAGELIRILQRALFMHFRGGCIHSRIGQPGGGNPLSAIPPSRSML"
gene            5948. .8761
                /gene="tat"
CDS             join(5948. .6162,8487. .8577)
                /gene="tat"
                /codon_start=1
                /product="tat"
                /protein_id="AAB99963.1"
                /db_xref="GI:2828042"
                /translation="MEPVDPRLEPWKHPGSKPKTACTNCYCKKCCFHCQVCFITKALG
                ISYGRKKRRQRRRAHQNSQTHQASLSKQPSSQPRGDPTGPKEQKKKVERETETDPVH"
gene            6087. .8761
                /gene="rev"
CDS             join(6087. .6162,8487. .8761)
                /gene="rev"
                /codon_start=1
                /product="rev"
                /protein_id="AAB99964.1"
                /db_xref="GI:2828043"
                /translation="MAGRSGDSDEELIRTVRLIKLLYQSNPPPSLEGTRQARRNRRRR
                WRERQRQIRSISERILGTYLGRSAEPVPLQLPPLERLTLDCNEDCGTSGTQGVGSPQI
                LVESPTVLESGTKE"
gene            6179. .6424
                /gene="vpu"
CDS             6179. .6424
                /gene="vpu"
                /codon_start=1
                /product="vpu"
                /protein_id="AAB99965.1"
                /db_xref="GI:2828044"
                /translation="MQPIQIAIVALVVAIIIAIVVWSIVIIEYRKILRQRKIDRLIDR
                LIERAEDSGNESEGEISALVEMGVEMGHHAPWDVDDL"
gene            6339. .8903
                /gene="env"
CDS             6339. .8903
                /gene="env"
                /codon_start=1
                /product="env"
                /protein_id="AAB99966.1"
                /db_xref="GI:2828045"
                /translation="MRVKEKYQHLWRWGWRWGTMLLGMLMICSATEKLWVTVYYGVPV
                WREAATTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEVVLGNVTENFNMWKNNMVD
                QMHEDIISLWDESLKPCVKLTPLCVTLNCTNLNITKNTTNPTSSSWGMMEKGEIKNCS
                FYITTSIRNKVKKEYALFNRLDVVPIENTNNTKYRLISCNTSVITQACPKVSFQPIPI
                HYCVPAGFAMLKCNNKTFNGSGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAEEDIVIR
                SENFTDNAKTIIVQLNESVVINCTRPNNNTRRRLSIGPGRAFYARRNIIGDIRQAHCN
                ISRAKWNNTLQQIVIKLREKFRNKTIAFNQSSGGDPEIVMHSFNCGGEFFYCNTAQLF
                NSTWNVTGGTNGTEGNDIITLQCRIKQIINMWQKVGKAMYAPPITGQIRCSSNITGLL
                LTRDGGNSTETETEIFRPGGGDMRDNWRSELYKYKVVRIEPIGVAPTRAKRRTVQREK
                RAVGIGAVFLGFLGAAGSTMGAASVTLTVQARLLLSGIVQQQNNLLRAIEAQQHMLQL
                TVWGIKQLQARVLALERYLRDQQLMGIWGCSGKLICTTSVPWNVSWSNKSVDDIWNNM
                TWMEWEREIDNYTDYIYDLLEKSQTQQEKNEKELLELDKWASLWNWFDITNWLWYIRL
                FIMIVGGLIGLRIVFAVLSIVNRVRQGYSPLSFQTLLPASRGPDRPEGTEEEGGERDR
                DRSGPLVNGSLALIWDDLRSLCLFSYHRLRDLLLIVTRIVELLGRRGWEALKYWWNLL
                QYWSQELKNSAVSLLNATAIAVAEGTDRVIEVVQGACRAIRHIPRRIRQGLERILL"
gene            8922. .9713
                /gene="nef"
CDS             8922. .9713
                /gene="nef"
                /codon_start=1
                /product="nef protein"
                /protein_id="AAB99967.1"
                /db_xref="GI:2828046"
                /translation="MGGAISMRRSRPSGDLRQRLLRARGETYGRLLGEVEDGYSQSPG
                GLDKGLSSLSCEGQKYNQGQYMNTPWRNPAEEREKLAYRKQNMDDIDEEDDDLVGVSV
                RPKVPLRTMSYKLAIDMSHFIKEKGGLEGIYYSARRHRILDIYLEKEEGIIPDWQDYT
                SGPGIRYPKTFGWLWKLVPVNVSDEAQEDEEHYLMHPAQTSQWDDPWGEVLAWKFDPT
                LAYTYEAYVRYPEEFGSKSGLSEEEVRRRLTARGLLNMADKKETR"
LTR             9307. .10000
                /note="3' long terminal repeat"
misc_feature    9307. .9823
                /note="U3 region"
repeat_region   9824. .10000
                /note="R region"
BASE COUNT     3466 a   1817 c   2500 g   2217 t
ORIGIN

  Query Match             100.0%;  Score 2564;  DB 14;  Length 10000;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2564;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2332 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 2391

Qy      61 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2392 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 2451

Qy     121 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2452 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 2511

Qy     181 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2512 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 2571

Qy     241 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2572 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 2631
```

```
Qy     301 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2632 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 2691

Qy     361 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2692 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 2751

Qy     421 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2752 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 2811

Qy     481 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2812 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 2871

Qy     541 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2872 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 2931

Qy     601 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2932 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 2991

Qy     661 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2992 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 3051

Qy     721 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3052 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 3111

Qy     781 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3112 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 3171

Qy     841 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3172 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 3231

Qy     901 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3232 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 3291

Qy     961 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3292 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 3351

Qy    1021 AAAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACA 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3352 AAAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACA 3411

Qy    1081 CATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATA 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3412 CATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATA 3471

Qy    1141 GTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAG 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3472 GTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAG 3531

Qy    1201 TGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCA 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3532 TGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCA 3591

Qy    1261 CCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTAT 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3592 CCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTAT 3651

Qy    1321 TATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGAT 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3652 TATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGAT 3711

Qy    1381 AGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAA 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3712 AGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAA 3771

Qy    1441 GCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAA 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3772 GCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAA 3831

Qy    1501 TATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAA 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3832 TATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAA 3891

Qy    1561 ATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAA 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3892 ATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAA 3951

Qy    1621 GGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTC 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3952 GGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTC 4011

Qy    1681 TTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAA 1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4012 TTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAA 4071

Qy    1741 GAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGT 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4072 GAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGT 4131

Qy    1801 GATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACT 1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4132 GATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACT 4191

Qy    1861 TGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCT 1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4192 TGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCT 4251

Qy    1921 AGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTT 1980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4252 AGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTT 4311

Qy    1981 CTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAAC 2040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4312 CTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAAC 4371

Qy    2041 TTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGG 2100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4372 TTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGG 4431

Qy    2101 GTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAAT 2160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4432 GTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAAT 4491

Qy    2161 CAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTT 2220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4492 CAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTT 4551

Qy    2221 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTA 2280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4552 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTA 4611

Qy    2281 ATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTT 2340
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4612 ATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTT 4671

Qy    2341 AAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAG 2400
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4672 AAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAG 4731

Qy    2401 CTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTA 2460
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4732 CTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTA 4791

Qy    2461 CCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGT 2520
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    4792 CCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGT 4851

Qy    2521 TCCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 2564
           ||||||||||||||||||||||||||||||||||||||||||||
Db    4852 TCCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 4895


RESULT 3
AF038399
LOCUS       AF038399               10006 bp    DNA     linear   VRL 02-FEB-1998
DEFINITION  Simian-Human immunodeficiency virus strain SHIV-4, complete genome.
ACCESSION   AF038399
VERSION     AF038399.1  GI:2828047
KEYWORDS    .
SOURCE      Simian-Human immunodeficiency virus.
  ORGANISM  Simian-Human immunodeficiency virus
            Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
            lentivirus group.
REFERENCE   1  (bases 1 to 10006)
  AUTHORS   Li,J., Lord,C.I., Haseltine,W., Letvin,N.L. and Sodroski,J.
  TITLE     Infection of cynomolgus monkeys with a chimeric HIV-1/SIVmac virus
            that expresses the HIV-1 envelope glycoproteins
  JOURNAL   J. Acquir. Immune Defic. Syndr. 5 (7), 639-646 (1992)
  MEDLINE   92309177
   PUBMED   1613662
REFERENCE   2  (bases 1 to 10006)
  AUTHORS   Reimann,K.A., Li,J.T., Voss,G., Lekutis,C., Tenner-Racz,K.,
            Racz,P., Lin,W., Montefiori,D.C., Lee-Parritz,D.E., Lu,Y.,
            Collman,R.G., Sodroski,J. and Letvin,N.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-DEC-1997) Human Retrovirology, Dana-Farber Cancer
            Institute, 44 Binney Street, Boston, MA 02115, USA
FEATURES             Location/Qualifiers
     source          1. .10006
                     /organism="Simian-Human immunodeficiency virus"
                     /proviral
                     /strain="SHIV-4; HXBc2"
                     /db_xref="taxon:57667"
     LTR             1. .301
                     /note="5' long terminal repeat"
     repeat_region   1. .177
                     /note="R region"
     misc_feature    178. .301
                     /note="U5 region"
     gene            536. .2068
                     /gene="gag"
     CDS             536. .2068
                     /gene="gag"
                     /codon_start=1
                     /product="gag"
                     /protein_id="AAB99968.1"
                     /db_xref="GI:2828048"
                     /translation="MGVRNSVLSGKKADELEKIRLRPNGKKKYMLKHVVWAANELDRF
                     GLAESLLENKEGCQKILSVLAPLVPTGSENLKSLYNTVCVIWCIHAEEKVKHTEEAKQ
                     IVQRHLVVETGTTETMPKTSRPTAPSSGRGGNYPVQQIGGNYVHLPLSPRTLNAWVKL
                     IEEKKFGAEVVPGFQALSEGCTPYDINQMLNCVGDHQAAMQIIRDIINEEAADWDLQH
                     PQPAPQQGQLREPSGSDIAGTTSSVDEQIQWMYRQQNPIPVGNIYRRWIQLGLQKCVR
                     MYNPTNILDVKQGPKEPFQSYVDRFYKSLRAEQTDAAVKNWMTQTLLIQNANPDCKLV
                     LKGLGVNPTLEEMLTACQGVGGPGQKARLMAEALKEALAPVPIPFAAAQQRGPRKPIK
                     CWNCGKEGHSARQCRAPRRQGCWKCGKMDHVMAKCPDRQAGFLGLGPWGKKPRNFPMA
                     QVHQGLMPTAPPEDPAVDLLKNYMQLGKQQREKQRESREKPYKEVTEDLLHLNSLFGG
                     DQ"
     gene            <1714. .4893
                     /gene="pol"
     CDS             <1714. .4893
                     /gene="pol"
                     /codon_start=1
                     /product="pol"
                     /protein_id="AAB99969.1"
                     /db_xref="GI:2828049"
                     /translation="LELWERGTLCKAMQSPKKTGMLEMWKNGPCYGQMPRQTGGFFRP
                     WSMGKEAPQFPHGSSASGADANCSPRGPSCGSAKELHAVGQAAERKAERKQREALQGG
                     DRGFAAPQFSLWRRPVVTAHIEGQPVEVLLDTGADDSIVTGIELGPHYTPKIVGGIGG
                     FINTKEYKNVEIEVLGKRIKGTIMTGDTPINIFGRNLLTALGMSLNFPIAKVEPVKVA
                     LKPGKDGPKLKQWPLSKEKIVALREICEKMEKDGQLEEAPPTNPYNTPTFAIKKKDKN
                     KWRMLIDFRELNRVTQDFTEVQLGIPHPAGLAKRKRITVLDIGDAYFSIPLDEEFRQY
                     TAFTLPSVNNAEPGKRYIYKVLPQGWKGSPAIFQYTMRHVLEPFRKANPDVTLVQYMD
                     DILIASDRTDLEHDRVVLQSKELLNSIGFSTPEEKFQKDPPFQWMGYELWPTKWKLQK
                     IELPQRETWTVNDIQKLVGVLNWAAQIYPGIKTKHLCRLIRGKMTLTEEVQWTEMAEA
                     EYEENKIILSQEQEGCYYQEGKPLEATVIKSQDNQWSYKIHQEDKILKVGKFAKIKNT
                     HTNGVRLLAHVIQKIGKEAIVIWGQVPKFHLPVEKDVWEQWWTDYWQVTWIPEWDFIS
                     TPPLVRLVFNLVKDPIEGEETYYTDGSCNKQSKEGKAGYITDRGKDKVKVLEQTTNQQ
                     AELEAFLMALTDSGPKANIIVDSQYVMGIITGCPTESESRLVNQIIEEMIKKSEIYVA
                     WVPAHKGIGGNQEIDHLVSQGIRQVLFLEKIEPAQEEHDKYHSNVKELVFKFGLPRIV
                     ARQIVDTCDKCHQKGEAIHGQANSDLGTWQMDCTHLEGKIIIVAVHVASGFIEAEVIP
                     QETGRQTALFLLKLAGRWPITHLHTDNGANFASQEVKMVAWWAGIEHTFGVPYNPQSQ
                     GVVEAMNHHLKNQIDRIREQANSVETIVLMAVHCMNFKRRGGIGDMTPAERLINMITT
                     EQEIQFQQSKNSKFKNFRVYYREGRDQLWKGPGELLWKGEGAVILKVGTDIKVVPRRK
                     AKIIKDYGGGKEVDSSSHMEDTGEAREVA"
gene            4823. .5467
                /gene="vif"
CDS             4823. .5467
                /gene="vif"
                /codon_start=1
                /product="vif"
                /protein_id="AAB99970.1"
                /db_xref="GI:2828050"
                /translation="MEEEKRWIAVPTWRIPERLERWHSLIKYLKYKTKDLQKVCYVPH
                FKVGWAWWTCSRVIFPLQEGSHLEVQGYWHLTPEKGWLSTYAVRITWYSKNFWTDVTP
                NYADILLHSTYFPCFTAGEVRRAIRGEQLLSCCRFPRAHKYQVPSLQYLALKVVSDVR
                SQGENPTWKQWRRDNRRGLRMAKQNSRGDKQRGGKPPTKGANFPGLAKVLGILA"
gene            5295. .5633
                /gene="vpx"
CDS             5295. .5633
                /gene="vpx"
                /codon_start=1
                /product="vpx"
                /protein_id="AAB99971.1"
                /db_xref="GI:2828051"
                /translation="MSDPRERIPPGNSGEETIGEAFEWLNRTVEEINREAVNHLPREL
                IFQVWQRSWEYWHDEQGMSPSYVKYRYLCLIQKALFMHCKKGCRCLGEGHGAGGWRPG
                PPPPPPPGLA"
gene            5634. .5939
                /gene="vpr"
CDS             5634. .5939
                /gene="vpr"
                /codon_start=1
                /product="vpr"
                /protein_id="AAB99972.1"
                /db_xref="GI:2828052"
                /translation="MEERPPENEGPQREPWDEWVVEVLEELKEEALKHFDPRLLTALG
                NHIYNRHGDTLEGAGELIRILQRALFMHFRGGCIHSRIGQPGGGNPLSAIPPSRSML"
gene            5948. .8580
                /gene="tat"
CDS             join(5948. .6162,8493. .8538)
                /gene="tat"
                /codon_start=1
                /product="tat"
                /protein_id="AAB99973.1"
                /db_xref="GI:2828053"
                /translation="MEPVDPRLEPWKHPGSKPKTACTNCYCKKCCFHCQVCFITKALG
                ISYGRKKRRQRRRAHQNSQTHQASLSKQPTSQPRGDPTGPKE"
gene            6087. .8767
                /gene="rev"
CDS             join(6087. .6162,8493. .8767)
                /gene="rev"
                /codon_start=1
                /product="rev"
                /protein_id="AAB99974.1"
                /db_xref="GI:2828054"
                /translation="MAGRSGDSDEELIRTVRLIKLLYQSNPPPNPEGTRQARRNRRRR
                WRERQRQIHSISERILGTYLGRSAEPVPLQLPPLERLTLDCNEDCGTSGTQGVGSPQI
                LVESPTVLESGTKE"
gene            6179. .6424
                /gene="vpu"
CDS             6179. .6424
```

```
                     /gene="vpu"
                     /codon_start=1
                     /product="vpu"
                     /protein_id="AAB99975.1"
                     /db_xref="GI:2828055"
                     /translation="MQPIQIAIVALVVAIIIAIVVWSIVIIEYRKILRQRKIDRLIDR
                     LIERAEDSGNESEGEISALVEMGVEMGHHAPWDVDDL"
     gene            6339. .8909
                     /gene="env"
     CDS             6339. .8909
                     /gene="env"
                     /codon_start=1
                     /product="env"
                     /protein_id="AAB99976.1"
                     /db_xref="GI:2828056"
                     /translation="MRVKEKYQHLWRWGWRWGTMLLGMLMICSATEKLWVTVYYGVPV
                     WKEATTTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEVVLVNVTENFNMWKNDMVE
                     QMHEDIISLWDQSLKPCVKLTPLCVSLKCTDLKNDTNTNSSSGRMIMEKGEIKNCSFN
                     ISTSIRGKVQKEYAFFYKLDIIPIDNDTTSYKLTSCNTSVITQACPKVSFEPIPIHYC
                     APAGFAILKCNNKTFNGTGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAEEEVVIRSVN
                     FTDNAKTIIVQLNTSVEINCTRPNNNTRKRIRIQRGPGRAFVTIGKIGNMRQAHCNIS
                     RAKWNNTLKQIASKLREQFGNNKTIIFKQSSGGDPEIVTHSFNCGGEFFYCNSTQLFN
                     STWFNSTWSTEGSNNTEGSDTITLPCRIKQIINMWQKVGKAMYAPPISGQIRCSSNIT
                     GLLLTRDGGNSNNESEIFRPGGGDMRDNWRSELYKYKVVKIEPLGVAPTKAKRRVVQR
                     EKRAVGIGALFLGFLGAAGSTMGAASMTLTVQARQLLSGIVQQQNNLLRAIEAQQHLL
                     QLTVWGIKQLQARILAVERYLKDQQLLGIWGCSGKLICTTAVPWNASWSNKSLEQIWN
                     HTTWMEWDREINNYTSLIHSLIEESQNQQEKNEQELLELDKWASLWNWFNITNWLWYI
                     KLFIMIVGGLVGLRIVFAVLSIVNRVRQGYSPLSFQTHLPTPRGPDRPEGIEEEGGER
                     DRDRSIRLVNGSLALIWDDLRSLCLFSYHRLRDLLLIVTRIVELLGRRGWEALKYWWN
                     LLQYWSQELKNSAVSLLNATAIAVAEGTDRVIEVVQGACRAIRHIPRRIRQGLERILL
                     "
     gene            8928. .9719
                     /gene="nef"
     CDS             8928. .9719
                     /gene="nef"
                     /codon_start=1
                     /product="nef protein"
                     /protein_id="AAB99977.1"
                     /db_xref="GI:2828057"
                     /translation="MGGAISMRRSRPSGDLRQRLLRARGETYGRLLGEVEDGYSQSPG
                     GLDKGLSSLSCEGQKYNQGQYMNTPWRNPAEEREKLAYRKQNMDDIDEEDDDLVGVSV
                     RPKVPLRTMSYKLAIDMSHFIKEKGGLEGIYYSARRHRILDIYLEKEEGIIPDWQDYT
                     SGPGIRYPKTFGWLWKLVPVNVSDEAQEDEEHYLMHPAQTSQWDDPWGEVLAWKFDPT
                     LAYTYEAYVRYPEEFGSKSGLSEEEVRRRLTARGLLNMADKKETR"
     LTR             9313. .10006
                     /note="3' long terminal repeat"
     misc_feature    9313. .9829
                     /note="U3 region"
     repeat_region   9830. .10006
                     /note="R region"
BASE COUNT     3468 a   1827 c   2499 g   2212 t
ORIGIN

  Query Match             100.0%;  Score 2564;  DB 14;  Length 10006;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2564;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2332 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 2391

Qy      61 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2392 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 2451

Qy     121 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2452 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 2511

Qy     181 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2512 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 2571

Qy     241 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2572 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 2631

Qy     301 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2632 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 2691

Qy     361 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2692 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 2751

Qy     421 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2752 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 2811

Qy     481 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2812 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 2871

Qy     541 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2872 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 2931

Qy     601 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2932 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 2991

Qy     661 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2992 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 3051

Qy     721 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3052 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 3111

Qy     781 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3112 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 3171

Qy     841 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3172 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 3231

Qy     901 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3232 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 3291

Qy     961 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3292 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 3351

Qy    1021 AAAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3352 AAAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACA 3411

Qy    1081 CATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATA 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3412 CATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATA 3471

Qy    1141 GTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3472 GTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAG 3531

Qy    1201 TGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCA 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3532 TGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCA 3591

Qy    1261 CCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTAT 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3592 CCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTAT 3651

Qy    1321 TATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGAT 1380
```

```
Db   3652 TATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGAT 3711

Qy   1381 AGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAA 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3712 AGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAA 3771

Qy   1441 GCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAA 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3772 GCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAA 3831

Qy   1501 TATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAA 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3832 TATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAA 3891

Qy   1561 ATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAA 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3892 ATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAA 3951

Qy   1621 GGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTC 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3952 GGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTC 4011

Qy   1681 TTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4012 TTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAA 4071

Qy   1741 GAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGT 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4072 GAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGT 4131

Qy   1801 GATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACT 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4132 GATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACT 4191

Qy   1861 TGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCT 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4192 TGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCT 4251

Qy   1921 AGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTT 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4252 AGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTT 4311

Qy   1981 CTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAAC 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4312 CTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAAC 4371

Qy   2041 TTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGG 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4372 TTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGG 4431

Qy   2101 GTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAAT 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4432 GTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAAT 4491

Qy   2161 CAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTT 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4492 CAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTT 4551

Qy   2221 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTA 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4552 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTA 4611

Qy   2281 ATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTT 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4612 ATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTT 4671

Qy   2341 AAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAG 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4672 AAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAG 4731

Qy   2401 CTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTA 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4732 CTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTA 4791

Qy   2461 CCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGT 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4792 CCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGT 4851

Qy   2521 TCCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 2564
          |||||||||||||||||||||||||||||||||||||||||||
Db   4852 TCCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 4895
```

```
RESULT 4
AF041850
LOCUS       AF041850               10006 bp    DNA     linear   VRL 03-FEB-1999
DEFINITION  Simian-Human immunodeficiency virus clone SHIV-HXBc2P 3.2, complete
            genome.
ACCESSION   AF041850
VERSION     AF041850.1  GI:2895564
KEYWORDS    .
SOURCE      Simian-Human immunodeficiency virus.
  ORGANISM  Simian-Human immunodeficiency virus
            Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
            lentivirus group.
REFERENCE   1  (bases 1 to 10006)
  AUTHORS   Cayabyab,M., Karlsson,G.B., Etemad-Moghadam,B.A., Hofmann,W.,
            Steenbeke,T., Halloran,M., Fanton,J.W., Axthelm,M.K., Letvin,N.L.
            and Sodroski,J.G.
  TITLE     Changes in human immunodeficiency virus type 1 envelope
            glycoproteins responsible for the pathogenicity of a multiply
            passaged simian-human immunodeficiency virus (SHIV-HXBc2)
  JOURNAL   J. Virol. 73 (2), 976-984 (1999)
  MEDLINE   99098984
   PUBMED   9882298
REFERENCE   2  (bases 1 to 10006)
  AUTHORS   Cayabyab,M., Karlsson,G.B., Etemad-Moghadam,B., Hofmann,W.,
            Halloran,M., Axthelm,M.W., Letvin,N.L. and Sodroski,J.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JAN-1998) Human Retrovirology, Dana-Farber Cancer
            Institute, 44 Binney Street, Boston, MA 02115, USA
FEATURES             Location/Qualifiers
     source          1. .10006
                     /organism="Simian-Human immunodeficiency virus"
                     /proviral
                     /db_xref="taxon:57667"
                     /clone="SHIV-HXBc2P 3.2"
                     /note="SHIV-HXBc2P3.2 is a molecularly cloned virus from
                     the pathogenic stock KU-1; Ku-1 was generated by multiply
                     passaging SHIV-4 (GenBank Accession Number AF038399)"
     LTR             1. .301
                     /note="5'LTR"
     repeat_region   1. .177
                     /note="R-region"
     repeat_region   178. .301
                     /note="U5-region"
     gene            536. .2068
                     /gene="gag"
     CDS             536. .2068
                     /gene="gag"
                     /codon_start=1
                     /product="gag protein"
                     /protein_id="AAD12134.1"
                     /db_xref="GI:2895565"
                     /translation="MGVRNSVLSGKKADELEKIRLRPNGKKKYMLKHVVWAANELDRF
                     GLAESLLENKEGCQKILSVLAPLVPTGSENLKSLYNTVCVIWCIHAEEKVKHTEEAKQ
                     IVQRHLVVETGTTETMPKTSRPTAPSSGRGGNYPVQQIGGNYVHLPLSPRTLNAWVKL
                     IEEKKFGAEVVPGFQALSEGCTPYDINQMLNCVGDHQAAMQIIRDIINEEAADWDLQH
                     PQPAPQQGQLREPSGSDIAGTTSSVDEQIQWMYRQQNPIPVGNIYRRWIQLGLQKCVR
                     MYNPTNILDVKQGPKEPFQSYVDRFYKSLRAEQTDAAVKNWMTQTLLIQNANPDCKLV
                     LKGLGVNPTLEEMLTACQGVGGPGQKARLMAEALKEALAPVPIPFAAAQQRGPRKPIK
                     CWNCGKEGHSARQCRAPRRQGCWKCGKMDHVMAKCPDRQAGFLGLGPWGKKPRNFPMA
                     QVHQGLMPTAPPEDPAVDLLKNYMQLGKQQREKQRESREKPYKEVTEDLLHLNSLFGG
                     DQ"
     gene            <1714. .4893
```

```
                     /gene="pol"
     CDS             <1714. .4893
                     /gene="pol"
                     /codon_start=1
                     /product="pol protein"
                     /protein_id="AAD12135.1"
                     /db_xref="GI:2895566"
                     /translation="LELWERGTLCKAMQSPKKTGMLEMWKNGPCYGQMPRQTGGFFRP
                     WSMGKEAPQFPHGSSASGADANCSPRGPSCGSAKELHAVGQAAERKAERKQREALQGG
                     DRGFAAPQFSLWRRPVVTAHIEGQPVEVLLDTGADDSIVTGIELGPHYTPKIVGGIGG
                     FINTKEYKNVEIEVLGKRIKGTIMTGDTPINIFGRNLLTALGMSLNFPIAKVEPVKVA
                     LKPGKDGPKLKQWPLSKEKIVALREICEKMEKDGQLEEAPPTNPYNTPTFAIKKKDKN
                     KWRMLIDFRELNRVTQDFTEVQLGIPHPAGLAKRKRITVLDIGDAYFSIPLDEEFRQY
                     TAFTLPSVNNAEPGKRYIYKVLPQGWKGSPAIFQYTMRHVLEPFRKANPDVTLVQYMD
                     DILIASDRTDLEHDRVVLQSKELLNSIGFSTPEEKFQKDPPFQWMGYELWPTKWKLQK
                     IELPQRETWTVNDIQKLVGVLNWAAQIYPGIKTKHLCRLIRGKMTLTEEVQWTEMAEA
                     EYEENKIILSQEQEGCYYQEGKPLEATVIKSQDNQWSYKIHQEDKILKVGKFAKIKNT
                     HTNGVRLLAHVIQKIGKEAIVIWGQVPKFHLPVEKDVWEQWWTDYWQVTWIPEWDFIS
                     TPPLVRLVFNLVKDPIEGEETYYTDGSCNKQSKEGKAGYITDRGKDKVKVLEQTTNQQ
                     AELEAFLMALTDSGPKANIIVDSQYVMGIITGCPTESESRLVNQIIEEMIKKSEIYVA
                     WVPAHKGIGGNQEIDHLVSQGIRQVLFLEKIEPAQEEHDKYHSNVKELVFKFGLPRIV
                     ARQIVDTCDKCHQKGEAIHGQANSDLGTWQMDCTHLEGKIIIVAVHVASGFIEAEVIP
                     QETGRQTALFLLKLAGRWPITHLHTDNGANFASQEVKMVAWWAGIEHTFGVPYNPQSQ
                     GVVEAMNHHLKNQIDRIREQANSVETIVLMAVHCMNFKRRGGIGDMTPAERLINMITT
                     EQEIQFQQSKNSKFKNFRVYYREGRDQLWKGPGELLWKGEGAVILKVGTDIKVVPRRK
                     AKIIKDYGGGKEVDSSSHMEDTGEAREVA"
     gene            4823. .5467
                     /gene="vif"
     CDS             4823. .5467
                     /gene="vif"
                     /codon_start=1
                     /product="vif protein"
                     /protein_id="AAD12136.1"
                     /db_xref="GI:2895567"
                     /translation="MEEEKRWIAVPTWRIPERLERWHSLIKYLKYKTKDLQKVCYVPH
                     FKVGWAWWTCSRVIFPLQEGSHLEVQGYWHLTPEKGWLSTYAVRITWYSKNFWTDVTP
                     NYADILLHSTYFPCFTAGEVRRAIRGEQLLSCCRFPRAHKYQVPSLQYLALKVVSDVR
                     SQGENPTWKQWRRDNRRGLRMAKQNSRGDKQRGGKPPTKGANFPGLAKVLGILA"
     gene            5295. .5633
                     /gene="vpx"
     CDS             5295. .5633
                     /gene="vpx"
                     /codon_start=1
                     /product="vpx protein"
                     /protein_id="AAD12137.1"
                     /db_xref="GI:2895568"
                     /translation="MSDPRERIPPGNSGEETIGEAFEWLNRTVEEINREAVNHLPREL
                     IFQVWQRSWEYWHDEQGMSPSYVKYRYLCLIQKALFMHCKKGCRCLGEGHGAGGWRPG
                     PPPPPPPGLA"
     gene            5634. .5939
                     /gene="vpr"
     CDS             5634. .5939
                     /gene="vpr"
                     /codon_start=1
                     /product="vpr protein"
                     /protein_id="AAD12138.1"
                     /db_xref="GI:2895569"
                     /translation="MEERPPENEGPQREPWDEWVVEVLEELKEEALKHFDPRLLTALG
                     NHIYNRHGDTLEGAGELIRILQRALFMHFRGGCIHSRIGQPGGGNPLSAIPPSRSML"
     gene            5948. .8538
                     /gene="tat"
     CDS             join(5948. .6162,8493. .8538)
                     /gene="tat"
                     /codon_start=1
                     /product="tat protein"
                     /protein_id="AAD12139.1"
                     /db_xref="GI:2895570"
                     /translation="MEPVDPRLEPWKHPGSKPKTACTNCYCKKCCFHCQVCFITKALG
                     ISYGRKKRRQRRRAHQNSQTHQASLSKQPTSQPRGDPTGPKE"
     gene            6087. .8767
                     /gene="rev"
     CDS             join(6087. .6162,8493. .8767)
                     /gene="rev"
                     /codon_start=1
                     /product="rev protein"
                     /protein_id="AAD12140.1"
                     /db_xref="GI:2895571"
                     /translation="MAGRSGDSDEELIRTVRLIKLLYQSNPPPNPEGTRQARRNRRRR
                     WRERQRQIHSISERILGTYLGRSAEPVPLQLPPLERLTLDCNEDCGTSGTQGVGSPQI
                     LVESPTVLESGTKE"
     gene            6179. .6424
                     /gene="vpu"
     CDS             6179. .6424
                     /gene="vpu"
                     /codon_start=1
                     /product="vpu protein"
                     /protein_id="AAD12141.1"
                     /db_xref="GI:2895572"
                     /translation="MQPIQIAIVALVVAIIIAIVVWSIVIIEYRKILRQRKIDRLIDR
                     LIERAEDSGNESEGEISALVEMGVEMGHHAPWDVDDL"
     gene            6339. .8909
                     /gene="env"
     CDS             6339. .8909
                     /gene="env"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAD12142.1"
                     /db_xref="GI:2895573"
                     /translation="MRVKEKYQHLWRWGWRWGTMLLGMLMICSATEKLWVTVYYGVPV
                     WKEATTTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEVVLVNVTENFNMWKNDMVE
                     QMHEDIISLWDQSLKPCVKLTPLCVSLNCTDLKNDTNTNSSSGRMIMEKGEIKNCSFN
                     ISTSIRGKVQKEYAFFYKLDIIPIDNDTTSYTLTSCNTSVISQACPKVSFEPIPIHYC
                     APAGFAILKCNNKTFNGTGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAEEEVVIRSVN
                     FMDNAKTIIVQLNTSVEINCTRPSNNTIKRIRIQRGPGRAFVTMGKIGDMRQAHCNIS
                     RAKWNNTLKQIASKLREQFGNNKTIIFKQSSGGDPEIVTHSFNCGGEFFYCNSTQLFN
                     STWFNSTWSTEGSNNTEGSDTITLPCRIKQIINMWQKVGKAMYAPPISGQIRCSSNIT
                     GLLLTRDGGKGNNESEIFRPGGGDMRDNWRSELYKYKVVKIEPLGVAPTKAKRRVVQR
                     EKRAVGIGALFLGFLGAAGSTMGAASMTLTVQARQLLSGIVQQQNNLLRAIEAQQHLL
                     QLTVWGIKQLQARILAVERYLKDQQLLGIWGCSGKLICTTAVPWNASWSNKSLEQIWN
                     HMTWMEWDREINNYTSLIHSLIEESQNQQEKNEQELLELDKWASLWNWFDITNWLWYI
                     KLFIMIVGGLVGLRIVFAVLSIVNRVRQGYSPLSFQTHLPTPRGPDRPEGIEEEGGER
                     DRDRSIRLVNGSLALIWDDLRSLCLFSYHRLRDLLLIVTRIVELLGRRGWEALKYWWN
                     LLQYWSQELKNSAVSLLNATAIAVAEGTDRVIEVVQGACRAIRHIPRRIRQGLERILL
                     "
     gene            8928. .9719
                     /gene="nef"
     CDS             8928. .9719
                     /gene="nef"
                     /codon_start=1
                     /product="nef protein"
                     /protein_id="AAD12143.1"
                     /db_xref="GI:2895574"
                     /translation="MGGAISMRRSRPSGDLRQRLLRARGETYGRLLGEVEDGYSQSPG
                     GLDKGLSSLSCEGQKYNQGQYMNTPWRNPAEEREKLAYRKQNMDDIDEEDDDLVGVSV
                     RPKVPLRTMSYKLAIDMSHFIKEKGGLEGIYYSARRHRILDIYLEKEEGIIPDWQDYT
                     SGPGIRYPKTFGWLWKLVPVNVSDEAQEDEEHYLMHPAQTSQWDDPWGEVLAWKFDPT
                     LAYTYEAYVRYPEEFGSKSGLSEEEVRRRLTARGLLNMADKKETR"
     LTR             9313. .10006
                     /note="5'LTR"
     repeat_region   9313. .9829
                     /note="U3-region"
     repeat_region   9830. .10006
                     /note="R-region"
BASE COUNT     3459 a   1824 c   2505 g   2218 t
ORIGIN

  Query Match           100.0%;  Score 2564;  DB 14;  Length 10006;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2564;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2332 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 2391

Qy     61 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   2392 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 2451

Qy    121 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2452 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 2511

Qy    181 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2512 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 2571

Qy    241 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2572 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 2631

Qy    301 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2632 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 2691

Qy    361 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2692 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 2751

Qy    421 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2752 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 2811

Qy    481 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2812 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 2871

Qy    541 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2872 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 2931

Qy    601 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2932 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 2991

Qy    661 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2992 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 3051

Qy    721 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3052 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 3111

Qy    781 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3112 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 3171

Qy    841 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3172 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 3231

Qy    901 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3232 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 3291

Qy    961 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3292 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 3351

Qy   1021 AAAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3352 AAAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACA 3411

Qy   1081 CATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATA 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3412 CATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATA 3471

Qy   1141 GTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3472 GTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAG 3531

Qy   1201 TGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCA 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3532 TGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCA 3591

Qy   1261 CCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTAT 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3592 CCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTAT 3651

Qy   1321 TATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGAT 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3652 TATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGAT 3711

Qy   1381 AGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAA 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3712 AGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAA 3771

Qy   1441 GCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAA 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3772 GCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAA 3831

Qy   1501 TATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAA 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3832 TATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAA 3891

Qy   1561 ATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAA 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3892 ATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAA 3951

Qy   1621 GGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTC 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3952 GGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTC 4011

Qy   1681 TTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4012 TTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAA 4071

Qy   1741 GAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGT 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4072 GAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGT 4131

Qy   1801 GATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACT 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4132 GATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACT 4191

Qy   1861 TGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCT 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4192 TGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCT 4251

Qy   1921 AGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTT 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4252 AGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTT 4311

Qy   1981 CTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAAC 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4312 CTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAAC 4371

Qy   2041 TTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGG 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4372 TTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGG 4431

Qy   2101 GTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAAT 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4432 GTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAAT 4491

Qy   2161 CAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTT 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4492 CAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTT 4551

Qy   2221 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTA 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4552 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTA 4611
```

```
Qy    2281 ATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTT 2340
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4612 ATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTT 4671

Qy    2341 AAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAG 2400
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4672 AAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAG 4731

Qy    2401 CTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTA 2460
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4732 CTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTA 4791

Qy    2461 CCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGT 2520
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4792 CCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGT 4851

Qy    2521 TCCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 2564
           |||||||||||||||||||||||||||||||||||||||||||
Db    4852 TCCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 4895


RESULT 5
AY033146
LOCUS       AY033146               10279 bp    DNA     linear   VRL 05-JUL-2001
DEFINITION  Simian immunodeficiency virus clone 17E-Fr, complete sequence.
ACCESSION   AY033146
VERSION     AY033146.1  GI:14600336
KEYWORDS    .
SOURCE      Simian immunodeficiency virus 17E-Fr.
  ORGANISM  Simian immunodeficiency virus 17E-Fr
            Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
            lentivirus group.
REFERENCE   1  (bases 1 to 10279)
  AUTHORS   Flaherty,M.T., Hauer,D.A., Mankowski,J.L., Zink,M.C. and
            Clements,J.E.
  TITLE     Molecular and biological characterization of a neurovirulent
            molecular clone of simian immunodeficiency virus
  JOURNAL   J. Virol. 71 (8), 5790-5798 (1997)
  MEDLINE   97366637
   PUBMED   9223467
REFERENCE   2  (bases 1 to 10279)
  AUTHORS   Flaherty,M.T., Hauer,D.A., Mankowski,J.L., Zink,M.C. and
            Clements,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-APR-2001) Comp. Med., Johns Hopkins U. Sch. of Med.,
            600 N. Wolfe St., Baltimore, MD 21287, USA
FEATURES             Location/Qualifiers
     source          1. .10279
                     /organism="Simian immunodeficiency virus 17E-Fr"
                     /proviral
                     /db_xref="taxon:160753"
                     /clone="17E-Fr"
                     /note="recombinant of SIV/17E-Br and SIVmac239; envelope,
                     nef, and 3' LTR from SIV/17E-Br swarm were cloned in place
                     of the same genes in an infectious molecular clone of
                     SIVmac239"
     LTR             1. .818
                     /note="5' long terminal repeat"
     misc_feature    1. .517
                     /note="U3 region"
     misc_feature    518. .694
                     /note="R region"
     misc_feature    695. .818
                     /note="U5 region"
     gene            1053. .2585
                     /gene="gag"
     CDS             1053. .2585
                     /gene="gag"
                     /codon_start=1
                     /product="gag polyprotein"
                     /protein_id="AAK64193.1"
                     /db_xref="GI:14600340"
                     /translation="MGVRNSVLSGKKADELEKIRLRPNGKKKYMLKHVVWAANELDRF
                     GLAESLLENKEGCQKILSVLAPLVPTGSENLKSLYNTVCVIWCIHAEEKVKHTEEAKQ
                     IVQRHLVVETGTTETMPKTSRPTAPSSGRGGNYPVQQIGGNYVHLPLSPRTLNAWVKL
                     IEEKKFGAEVVPGFQALSEGCTPYDINQMLNCVGDHQAAMQIIRDIINEEAADWDLQH
                     PQPAPQQGQLREPSGSDIAGTTSSVDEQIQWMYRQQNPIPVGNIYRRWIQLGLQKCVR
                     MYNPTNILDVKQGPKEPFQSYVDRFYKSLRAEQTDAAVKNWMTQTLLIQNANPDCKLV
                     LKGLGVNPTLEEMLTACQGVGGPGQKARLMAEALKEALAPVPIPFAAAQQRGPRKPIK
                     CWNCGKEGHSARQCRAPRRQGCWKCGKMDHVMAKCPDRQAGFLGLGPWGKKPRNFPMA
                     QVHQGLMPTAPPEDPAVDLLKNYMQLGKQQREKQRESREKPYKEVTEDLLHLNSLFGG
                     DQ"
     gene            <2228. .5410
                     /gene="pol"
     CDS             <2228. .5410
                     /gene="pol"
                     /codon_start=1
                     /product="pol polyprotein"
                     /protein_id="AAK64194.1"
                     /db_xref="GI:14600341"
                     /translation="VLELWERGTLCKAMQSPKKTGMLEMWKNGPCYGQMPRQTGGFFR
                     PWSMGKEAPQFPHGSSASGADANCSPRGPSCGSAKELHAVGQAAERKAERKQREALQG
                     GDRGFAAPQFSLWRRPVVTAHIEGQPVEVLLDTGADDSIVTGIELGPHYTPKIVGGIG
                     GFINTKEYKNVEIEVLGKRIKGTIMTGDTPINIFGRNLLTALGMSLNFPIAKVEPVKV
                     ALKPGKDGPKLKQWPLSKEKIVALREICEKMEKDGQLEEAPPTNPYNTPTFAIKKKDK
                     NKWRMLIDFRELNRVTQDFTEVQLGIPHPAGLAKRKRITVLDIGDAYFSIPLDEEFRQ
                     YTAFTLPSVNNAEPGKRYIYKVLPQGWKGSPAIFQYTMRHVLEPFRKANPDVTLVQYM
                     DDILIASDRTDLEHDRVVLQSKELLNSIGFSTPEEKFQKDPPFQWMGYELWPTKWKLQ
                     KIELPQRETWTVNDIQKLVGVLNWAAQIYPGIKTKHLCRLIRGKMTLTEEVQWTEMAE
                     AEYEENKIILSQEQEGCYYQEGKPLEATVIKSQDNQWSYKIHQEDKILKVGKFAKIKN
                     THTNGVRLLAHVIQKIGKEAIVIWGQVPKFHLPVEKDVWEQWWTDYWQVTWIPEWDFI
                     STPPLVRLVFNLVKDPIEGEETYYTDGSCNKQSKEGKAGYITDRGKDKVKVLEQTTNQ
                     QAELEAFLMALTDSGPKANIIVDSQYVMGIITGCPTESESRLVNQIIEEMIKKSEIYV
                     AWVPAHKGIGGNQEIDHLVSQGIRQVLFLEKIEPAQEEHDKYHSNVKELVFKFGLPRI
                     VARQIVDTCDKCHQKGEAIHGQANSDLGTWQMDCTHLEGKIIIVAVHVASGFIEAEVI
                     PQETGRQTALFLLKLAGRWPITHLHTDNGANFASQEVKMVAWWAGIEHTFGVPYNPQS
                     QGVVEAMNHHLKNQIDRIREQANSVETIVLMAVHCMNFKRRGGIGDMTPAERLINMIT
                     TEQEIQFQQSKNSKFKNFRVYYREGRDQLWKGPGELLWKGEGAVILKVGTDIKVVPRR
                     KAKIIKDYGGGKEVDSSSHMEDTGEAREVA"
     misc_feature    5084. .5099
                     /gene="pol"
                     /note="polypurine tract"
     gene            5340. .5984
                     /gene="vif"
     CDS             5340. .5984
                     /gene="vif"
                     /codon_start=1
                     /product="vif protein"
                     /protein_id="AAK64195.1"
                     /db_xref="GI:14600342"
                     /translation="MEEEKRWIAVPTWRIPERLERWHSLIKYLKYKTKDLQKVCYVPH
                     FKVGWAWWTCSRVIFPLQEGSHLEVQGYWHLTPEKGWLSTYAVRITWYSKNFWTDVTP
                     NYADILLHSTYFPCFTAGEVRRAIRGEQLLSCCRFPRAHKYQVPSLQYLALKVVSDVR
                     SQGENPTWKQWRRDNRRGLRMAKQNSRGDKQRGGKPPTKGANFPGLAKVLGILA"
     gene            5812. .6150
                     /gene="vpx"
     CDS             5812. .6150
                     /gene="vpx"
                     /codon_start=1
                     /product="vpx protein"
                     /protein_id="AAK64196.1"
                     /db_xref="GI:14600343"
                     /translation="MSDPRERIPPGNSGEETIGEAFEWLNRTVEEINREAVNHLPREL
                     IFQVWQRSWEYWHDEQGMSPSYVKYRYLCLIQKALFMHCKKGCRCLGEGHGAGGWRPG
                     PPPPPPPGLA"
     gene            6151. .6456
                     /gene="vpr"
     CDS             6151. .6456
                     /gene="vpr"
                     /codon_start=1
                     /product="vpr protein"
                     /protein_id="AAK64197.1"
                     /db_xref="GI:14600344"
                     /translation="MEERPPENEGPQREPWDEWVVEVLEELKEEALKHFDPRLLTALG
                     NHIYNRHGDTLEGAGELIRILQRALFMHFRGGCIHSRIGQPGGGNPLSAIPPSRSML"
     gene            6302. .8902
```

```
                     /gene="tat"
     CDS             join(6302. .6597,8803. .8902)
                     /gene="tat"
                     /note="transactivator of viral transcription"
                     /codon_start=1
                     /product="tat protein"
                     /protein_id="AAK64190.1"
                     /db_xref="GI:14600337"
                     /translation="METPLREQENSLESSNERSSCISEADASTPESANLGEEILSQLY
                     RPLEACYNTCYCKKCCYHCQFCFLKKGLGICYEQSRKRRRTPKKAKANTSSASNNRPI
                     SNRTRHCQPEKAKKETVEKAVATAPGLDR"
     gene            6528. .9059
                     /gene="rev"
     CDS             join(6528. .6597,8803. .9059)
                     /gene="rev"
                     /codon_start=1
                     /product="rev protein"
                     /protein_id="AAK64191.1"
                     /db_xref="GI:14600338"
                     /translation="MNNHEREEELRKRLRLIHLLHQTTDPYPTGPGTANQRRQRRRRW
                     RRRWQQLLALTDRIYSFPDPPTDTPLDLAIQQLQNLAIESIPDPPTNTPEALCDPTED
                     SRSPQD"
     gene            6604. .9243
                     /gene="env"
                     /note="contains premature stop codon"
     CDS             6604. .8895
                     /gene="env"
                     /codon_start=1
                     /product="truncated envelope glycoprotein"
                     /protein_id="AAK64192.1"
                     /db_xref="GI:14600339"
                     /translation="MGCLGNQLLIAILLLSVYGIYCTLYVTVFYGVPVWRNATIPLFC
                     ATKNRDTWGTTQCLPDNGDYSEMALNVTESFDAWNNTVTEQAIEDVWQLFETSIKPCV
                     KLSPLCITMRCNKSETDRWGLTKSITTTASTTSTTASARVDMVNETSSCIAQDNCAGL
                     EQEQMISCKFNMTGLNRDKKKEYNETWYSADLVCEQGNNTGNESRCYMNHCNTSVIKE
                     SCDKHYWDAIRFRYCAPPGYALLRCNDTNYSGFMPKCSKVVVSSCTRMMETQTSTWFG
                     FNGTRAENRTYIYWHGRDNRTIISLNKYYNLTIKCRRPGNKTVLPVTIMSGLVFHSQR
                     INDRPRQAWCWFGGKWKDAIKEVKQTIVKHPRYTGTNNTDKINLTAPRGGDPEVTFMW
                     TNCRGEFLYCKMNWFLNWVEDRNTANQKPKEQHKRNYVPCHIRQIINTWHKVGKNVYL
                     PPREGDLTCNSTVTSLIANIDWIDGNQTNITMSAEVAELYRLELGDYKLVEITPIGLA
                     PTDVKRYTTGGTSRNKRGVFVLGFLGFLATAGSAMGAASLTLTAQSRTLLAGIVQQQQ
                     QLLDVVKRQQELLRLTVWGTKNLQTRVTAIEKYLKDQAQLNTWGCAFRQVCHTTVPWP
                     NASLIPKWNNETWQEWERKVDFLEENITALLEEAQIQQEKNMYELQKLNSWNVFGNWF
                     DLASWIKYIQYGVYIVVGVILLRIVIYIVQVLAKLRQGYRPVFSSPPSYFQQTHIQQD
                     PALPTREGKEGDGGEGGGNSSWP"
     gene            9077. .9868
                     /gene="nef"
     CDS             9077. .9868
                     /gene="nef"
                     /codon_start=1
                     /product="nef protein"
                     /protein_id="AAK64198.1"
                     /db_xref="GI:14600345"
                     /translation="MGGAISMRRSRSSGDLRQRLLRARGETYGRLLGEVEDGYSQSPG
                     GLDKGLSSLSCEGQKYNQGQYMNTPWRNPAKEREKLAYRKQNMDDIDEQDDDLVGVSV
                     RPKVPLRTMSYKLAIDISHFIKEKGGLEGIYYSARRHRILDIYLEKEKGIIPDWQDYT
                     SGPGIRFPKTFGWLWKLVPVNVSDEAQEDEEHYLMHPAQTSQWDDPWGEVLAWKFDPT
                     LAYTYEAYVRYPEEFGSKSGLSEEEVRRRLTARGLLNMADKKETR"
     misc_feature    9445. .9460
                     /gene="nef"
                     /note="polypurine tract"
     LTR             9462. .10279
                     /note="3' long terminal repeat"
     misc_feature    9462. .9978
                     /note="U3 region"
     misc_feature    9979. .10155
                     /note="R region"
     polyA_signal    10132. .10137
     misc_feature    10156. .10279
                     /note="U5 region"
BASE COUNT     3470 a   1932 c   2566 g   2311 t
ORIGIN
```

```
  Query Match           100.0%;  Score 2564;  DB 14;  Length 10279;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 2564;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2849 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 2908

Qy      61 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2909 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 2968

Qy     121 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2969 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 3028

Qy     181 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3029 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 3088

Qy     241 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3089 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 3148

Qy     301 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3149 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 3208

Qy     361 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3209 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 3268

Qy     421 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3269 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 3328

Qy     481 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3329 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 3388

Qy     541 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3389 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 3448

Qy     601 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3449 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 3508

Qy     661 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3509 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 3568

Qy     721 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3569 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 3628

Qy     781 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3629 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 3688

Qy     841 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3689 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 3748

Qy     901 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3749 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 3808

Qy     961 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3809 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 3868

Qy    1021 AAAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACA 1080
```

```
Db   3869 AAAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACA 3928

Qy   1081 CATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATA 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3929 CATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATA 3988

Qy   1141 GTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3989 GTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAG 4048

Qy   1201 TGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCA 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4049 TGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCA 4108

Qy   1261 CCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTAT 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4109 CCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTAT 4168

Qy   1321 TATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGAT 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4169 TATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGAT 4228

Qy   1381 AGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAA 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4229 AGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAA 4288

Qy   1441 GCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAA 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4289 GCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAA 4348

Qy   1501 TATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAA 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4349 TATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAA 4408

Qy   1561 ATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAA 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4409 ATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAA 4468

Qy   1621 GGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTC 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4469 GGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTC 4528

Qy   1681 TTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4529 TTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAA 4588

Qy   1741 GAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGT 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4589 GAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGT 4648

Qy   1801 GATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACT 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4649 GATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACT 4708

Qy   1861 TGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCT 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4709 TGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCT 4768

Qy   1921 AGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTT 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4769 AGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTT 4828

Qy   1981 CTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAAC 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4829 CTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAAC 4888

Qy   2041 TTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGG 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4889 TTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGG 4948

Qy   2101 GTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAAT 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   4949 GTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAAT 5008

Qy   2161 CAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTT 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5009 CAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTT 5068

Qy   2221 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTA 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5069 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTA 5128

Qy   2281 ATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTT 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5129 ATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTT 5188

Qy   2341 AAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAG 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5189 AAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAG 5248

Qy   2401 CTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTA 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5249 CTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTA 5308

Qy   2461 CCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGT 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5309 CCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGT 5368

Qy   2521 TCCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 2564
          |||||||||||||||||||||||||||||||||||||||||||
Db   5369 TCCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 5412
```

```
RESULT 6
AY033233
LOCUS       AY033233               10279 bp    DNA     linear   VRL 05-JUL-2001
DEFINITION  Simian immunodeficiency virus clone 17E-Cl, complete sequence.
ACCESSION   AY033233
VERSION     AY033233.1  GI:14600346
KEYWORDS    .
SOURCE      Simian immunodeficiency virus 17E-Cl.
  ORGANISM  Simian immunodeficiency virus 17E-Cl
            Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
            lentivirus group.
REFERENCE   1  (bases 1 to 10279)
  AUTHORS   Anderson,M.G., Hauer,D., Sharma,D.P., Joag,S.V., Narayan,O.,
            Zink,M.C. and Clements,J.E.
  TITLE     Analysis of envelope changes acquired by SIVmac239 during
            neuroadaption in rhesus macaques
  JOURNAL   Virology 195 (2), 616-626 (1993)
  MEDLINE   93331720
   PUBMED   8337835
REFERENCE   2  (bases 1 to 10279)
  AUTHORS   Anderson,M.G., Hauer,D.A., Sharma,D.P., Joag,S.V., Narayan,O.,
            Zink,M.C. and Clements,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-APR-2001) Comp. Med., Johns Hopkins U. Sch. of Med.,
            600 N. Wolfe St., Baltimore, MD 21287, USA
FEATURES             Location/Qualifiers
     source          1. .10279
                     /organism="Simian immunodeficiency virus 17E-Cl"
                     /proviral
                     /db_xref="taxon:160754"
                     /clone="17E-Cl"
                     /note="recombinant of SIV/17E-Br and SIVmac239; SU portion
                     of envelope from SIV/17E-Br swarm was cloned in place of
                     the same gene in an infectious molecular clone of
                     SIVmac239"
     LTR             1. .818
                     /note="5' long terminal repeat"
     misc_feature    1. .517
                     /note="U3 region"
     misc_feature    518. .694
                     /note="R region"
     misc_feature    695. .818
```

```
                     /note="U5 region"
gene                 1053. .2585
                     /gene="gag"
CDS                  1053. .2585
                     /gene="gag"
                     /codon_start=1
                     /product="gag polyprotein"
                     /protein_id="AAK64202.1"
                     /db_xref="GI:14600350"
                     /translation="MGVRNSVLSGKKADELEKIRLRPNGKKKYMLKHVVWAANELDRF
                     GLAESLLENKEGCQKILSVLAPLVPTGSENLKSLYNTVCVIWCIHAEEKVKHTEEAKQ
                     IVQRHLVVETGTTETMPKTSRPTAPSSGRGGNYPVQQIGGNYVHLPLSPRTLNAWVKL
                     IEEKKFGAEVVPGFQALSEGCTPYDINQMLNCVGDHQAAMQIIRDIINEEAADWDLQH
                     PQPAPQQGQLREPSGSDIAGTTSSVDEQIQWMYRQQNPIPVGNIYRRWIQLGLQKCVR
                     MYNPTNILDVKQGPKEPFQSYVDRFYKSLRAEQTDAAVKNWMTQTLLIQNANPDCKLV
                     LKGLGVNPTLEEMLTACQGVGGPGQKARLMAEALKEALAPVPIPFAAAQQRGPRKPIK
                     CWNCGKEGHSARQCRAPRRQGCWKCGKMDHVMAKCPDRQAGFLGLGPWGKKPRNFPMA
                     QVHQGLMPTAPPEDPAVDLLKNYMQLGKQQREKQRESREKPYKEVTEDLLHLNSLFGG
                     DQ"
gene                 <2228. .5410
                     /gene="pol"
CDS                  <2228. .5410
                     /gene="pol"
                     /codon_start=1
                     /product="pol polyprotein"
                     /protein_id="AAK64203.1"
                     /db_xref="GI:14600351"
                     /translation="VLELWERGTLCKAMQSPKKTGMLEMWKNGPCYGQMPRQTGGFFR
                     PWSMGKEAPQFPHGSSASGADANCSPRGPSCGSAKELHAVGQAAERKAERKQREALQG
                     GDRGFAAPQFSLWRRPVVTAHIEGQPVEVLLDTGADDSIVTGIELGPHYTPKIVGGIG
                     GFINTKEYKNVEIEVLGKRIKGTIMTGDTPINIFGRNLLTALGMSLNFPIAKVEPVKV
                     ALKPGKDGPKLKQWPLSKEKIVALREICEKMEKDGQLEEAPPTNPYNTPTFAIKKDK
                     NKWRMLIDFRELNRVTQDFTEVQLGIPHPAGLAKRKRITVLDIGDAYFSIPLDEEFRQ
                     YTAFTLPSVNNAEPGKRYIYKVLPQGWKGSPAIFQYTMRHVLEPFRKANPDVTLVQYM
                     DDILIASDRTDLEHDRVVLQSKELLNSIGFSTPEEKFQKDPPFQWMGYELWPTKWKLQ
                     KIELPQRETWTVNDIQKLVGVLNWAAQIYPGIKTKHLCRLIRGKMTLTEEVQWTEMAE
                     AEYEENKIILSQEQEGCYYQEGKPLEATVIKSQDNQWSYKIHQEDKILKVGKFAKIKN
                     THTNGVRLLAHVIQKIGKEAIVIWGQVPKFHLPVEKDVWEQWWTDYWQVTWIPEWDFI
                     STPPLVRLVFNLVKDPIEGEETYYTDGSCNKQSKEGKAGYITDRGKDKVKVLEQTTNQ
                     QAELEAFLMALTDSGPKANIIVDSQYVMGIITGCPTESESRLVNQIIEEMIKKSEIYV
                     AWVPAHKGIGGNQEIDHLVSQGIRQVLFLEKIEPAQEEHDKYHSNVKELVFKFGLPRI
                     VARQIVDTCDKCHQKGEAIHGQANSDLGTWQMDCTHLEGKIIIVAVHVASGFIEAEVI
                     PQETGRQTALFLLKLAGRWPITHLHTDNGANFASQEVKMVAWWAGIEHTFGVPYNPQS
                     QGVVEAMNHHLKNQIDRIREQANSVETIVLMAVHCMNFKRRGGIGDMTPAERLINMIT
                     TEQEIQFQQSKNSKFKNFRVYYREGRDQLWKGPGELLWKGEGAVILKVGTDIKVVPRR
                     KAKIIKDYGGGKEVDSSSHMEDTGEAREVA"
misc_feature         5084. .5099
                     /gene="pol"
                     /note="polypurine tract"
gene                 5340. .5984
                     /gene="vif"
CDS                  5340. .5984
                     /gene="vif"
                     /codon_start=1
                     /product="vif protein"
                     /protein_id="AAK64204.1"
                     /db_xref="GI:14600352"
                     /translation="MEEEKRWIAVPTWRIPERLERWHSLIKYLKYKTKDLQKVCYVPH
                     FKVGWAWWTCSRVIFPLQEGSHLEVQGYWHLTPEKGWLSTYAVRITWYSKNFWTDVTP
                     NYADILLHSTYFPCFTAGEVRRAIRGEQLLSCCRFPRAHKYQVPSLQYLALKVVSDVR
                     SQGENPTWKQWRRDNRRGLRMAKQNSRGDKQRGGKPPTKGANFPGLAKVLGILA"
gene                 5812. .6150
                     /gene="vpx"
CDS                  5812. .6150
                     /gene="vpx"
                     /codon_start=1
                     /product="vpx protein"
                     /protein_id="AAK64205.1"
                     /db_xref="GI:14600353"
                     /translation="MSDPRERIPPGNSGEETIGEAFEWLNRTVEEINREAVNHLPREL
                     IFQVWQRSWEYWHDEQGMSPSYVKYRYLCLIQKALFMHCKKGCRCLGEGHGAGGWRPG
                     PPPPPPPGLA"
gene                 6151. .6456
                     /gene="vpr"
CDS                  6151. .6456
                     /gene="vpr"
                     /codon_start=1
                     /product="vpr protein"
                     /protein_id="AAK64206.1"
                     /db_xref="GI:14600354"
                     /translation="MEERPPENEGPQREPWDEWVVEVLEELKEEALKHFDPRLLTALG
                     NHIYNRHGDTLEGAGELIRILQRALFMHFRGGCIHSRIGQPGGGNPLSAIPPSRSML"
gene                 6302. .8902
                     /gene="tat"
CDS                  join(6302. .6597,8803. .8902)
                     /gene="tat"
                     /codon_start=1
                     /product="tat protein"
                     /protein_id="AAK64199.1"
                     /db_xref="GI:14600347"
                     /translation="METPLREQENSLESSNERSSCISEADASTPESANLGEEILSQLY
                     RPLEACYNTCYCKKCCYHCQFCFLKKGLGICYEQSRKRRRTPKKAKANTSSASNNRPI
                     SNRTRHCQPEKAKKETVEKAVATAPGLGR"
gene                 6528. .9059
                     /gene="rev"
CDS                  join(6528. .6597,8803. .9059)
                     /gene="rev"
                     /codon_start=1
                     /product="rev protein"
                     /protein_id="AAK64200.1"
                     /db_xref="GI:14600348"
                     /translation="MNNHEREEELRKRLRLIHLLHQTTDPYPTGPGTANQRRQRKRRW
                     RRRWQQLLALADRIYSFPDPPTDTPLDLAIQQLQNLAIESIPDPPTNTPEALCDPTED
                     SRSPQD"
gene                 6604. .9243
                     /gene="env"
CDS                  6604. .9243
                     /gene="env"
                     /note="envelope glycoprotein; includes both SU and TM
                     portions"
                     /codon_start=1
                     /product="env protein"
                     /protein_id="AAK64201.1"
                     /db_xref="GI:14600349"
                     /translation="MGCLGNQLLIAILLLSVYGIYCTLYVTVFYGVPVWRNATIPLFC
                     ATKNRDTWGTTQCLPDNGDYSEMALNVTESFDAWNNTVTEQAIEDVWQLFETSIKPCV
                     KLSPLCITMRCNKSETDRWGLTKSITTTASTTSTTASARVDMVNETSSCIAQDNCAGL
                     EQEQMISCKFNMTGLNRDKKKEYNETWYSADLVCEQGNNTGNESRCYMNHCNTSVIKE
                     SCDKHYWDAIRFRYCAPPGYALLRCNDTNYSGFMPKCSKVVVSSCTRMMETQTSTWFG
                     FNGTRAENRTYIYWHGRDNRTIISLNKYYNLTIKCRRPGNKTVLPVTIMSGLVFHSQR
                     INDRPRQAWCWFGGKWKDAIKEVKQTIVKHPRYTGTNNTDKINLTAPRGGDPEVTFMW
                     TNCRGEFLYCKMNWFLNWVEDRNTANQKPKEQHKRNYVPCHIRQIINTWHKVGKNVYL
                     PPREGDLTCNSTVTSLIANIDWIDGNQTNITMSAEVAELYRLELGDYKLVEITPIGLA
                     PTDVKRYTTGGTSRNKRGVFVLGFLGFLATAGSAMGAASLTLTAQSRTLLAGIVQQQQ
                     QLLDVVKRQQELLRLTVWGTKNLQTRVTAIEKYLKDQAQLNTWGCAFRQVCHTTVPWP
                     NASLIPKWNNETWQEWERKVDFLEENITALLEEAQIQQEKNMYELQKLNSWNVFGNWF
                     DLASWIKYIQYGVYIVVGVILLRIVIYIVQVLAKLRQGYRPVFSSPPSYFQQTHIQQD
                     PALPTREGKERDGGEGGGNSSWPWQIEYIHFLIRQLIRLLTWLFSNCRTLLSRVYQIL
                     QPILQRLSATLQRIREVLRTELTYLQYGWSYFHEAVQAVWRSATETLAGAWGDLWETL
                     RRGGRWILAIPRRIRQGLELTLL"
gene                 9077. .9868
                     /gene="nef"
                     /note="contains premature stop codon"
CDS                  9077. .9355
                     /gene="nef"
                     /codon_start=1
                     /product="truncated nef protein"
                     /protein_id="AAK64207.1"
                     /db_xref="GI:14600355"
                     /translation="MGGAISMRRSRPSGDLRQRLLRARGETYGRLLGEVEDGYSQSPG
                     GLDKGLSSLSCEGQKYNQGQYMNTPWRNPAEEREKLAYRKQNMDDIDE"
misc_feature         9445. .9460
                     /gene="nef"
                     /note="polypurine tract"
LTR                  9462. .10279
                     /note="3' long terminal repeat"
```

```
     misc_feature    9462. .9978
                     /note="U3 region"
     misc_feature    9979. .10155
                     /note="R region"
     polyA_signal    10132. .10137
     misc_feature    10156. .10279
                     /note="U5 region"
BASE COUNT     3468 a    1932 c    2569 g    2310 t
ORIGIN

  Query Match           100.0%;  Score 2564;  DB 14;  Length 10279;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 2564;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 60
Db    2849 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 2908

Qy      61 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 120
Db    2909 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 2968

Qy     121 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 180
Db    2969 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 3028

Qy     181 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 240
Db    3029 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 3088

Qy     241 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 300
Db    3089 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 3148

Qy     301 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 360
Db    3149 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 3208

Qy     361 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 420
Db    3209 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 3268

Qy     421 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 480
Db    3269 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 3328

Qy     481 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 540
Db    3329 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 3388

Qy     541 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 600
Db    3389 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 3448

Qy     601 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 660
Db    3449 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 3508

Qy     661 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 720
Db    3509 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 3568

Qy     721 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 780
Db    3569 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 3628

Qy     781 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 840
Db    3629 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 3688

Qy     841 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 900
Db    3689 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 3748

Qy     901 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 960
Db    3749 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 3808

Qy     961 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 1020
Db    3809 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 3868

Qy    1021 AAAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACA 1080
Db    3869 AAAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACA 3928

Qy    1081 CATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATA 1140
Db    3929 CATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATA 3988

Qy    1141 GTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAG 1200
Db    3989 GTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAG 4048

Qy    1201 TGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCA 1260
Db    4049 TGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCA 4108

Qy    1261 CCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTAT 1320
Db    4109 CCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTAT 4168

Qy    1321 TATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGAT 1380
Db    4169 TATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGAT 4228

Qy    1381 AGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAA 1440
Db    4229 AGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAA 4288

Qy    1441 GCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAA 1500
Db    4289 GCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAA 4348

Qy    1501 TATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAA 1560
Db    4349 TATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAA 4408

Qy    1561 ATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAA 1620
Db    4409 ATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAA 4468

Qy    1621 GGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTC 1680
Db    4469 GGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTC 4528

Qy    1681 TTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAA 1740
Db    4529 TTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAA 4588

Qy    1741 GAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGT 1800
Db    4589 GAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGT 4648

Qy    1801 GATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACT 1860
Db    4649 GATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACT 4708

Qy    1861 TGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCT 1920
Db    4709 TGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCT 4768

Qy    1921 AGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTT 1980
Db    4769 AGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTT 4828
```

```
Qy   1981 CTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAAC 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4829 CTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAAC 4888

Qy   2041 TTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGG 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4889 TTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGG 4948

Qy   2101 GTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAAT 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4949 GTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAAT 5008

Qy   2161 CAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTT 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5009 CAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTT 5068

Qy   2221 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTA 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5069 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTA 5128

Qy   2281 ATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTT 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5129 ATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTT 5188

Qy   2341 AAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAG 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5189 AAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAG 5248

Qy   2401 CTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTA 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5249 CTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTA 5308

Qy   2461 CCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGT 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5309 CCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGT 5368

Qy   2521 TCCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 2564
          |||||||||||||||||||||||||||||||||||||||||||
Db   5369 TCCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 5412
```

RESULT 7
SIVMM239

```
LOCUS       SIVMM239               10535 bp ss-RNA     linear   VRL 18-APR-2000
DEFINITION  Simian (macaque) immunodeficiency virus, isolate 239, complete
            proviral genome and flanking sequence.
ACCESSION   M33262
VERSION     M33262.1  GI:334647
KEYWORDS    .
SOURCE      Simian immunodeficiency virus.
  ORGANISM  Simian immunodeficiency virus
            Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
            lentivirus group.
REFERENCE   1  (bases 1 to 10535)
  AUTHORS   Kestler,H., Kodama,T., Ringler,D., Marthas,M., Pedersen,N.C.,
            Lackner,A., Regier,D., Sehgal,P., Daniel,M., King,N. and
            Desrosiers,R.
  JOURNAL   Unpublished
REFERENCE   2  (sites)
  AUTHORS   Kestler,H., Kodama,T., Ringler,D., Marthas,M., Pedersen,N.C.,
            Lackner,A., Regier,D., Sehgal,P., Daniel,M., King,N. and
            Desrosiers,R.
  TITLE     Induction of AIDS in rhesus monkeys by molecularly cloned simian
            immunodeficiency virus
  JOURNAL   Science 248 (4959), 1109-1112 (1990)
  MEDLINE   90260657
   PUBMED   2160735
COMMENT     [2]  sites; for [1].
            Kindly submitted in computer-readable form prior to publication by
            Dr. Desrosiers, New England Regional Primate Center, Harvard.  In
            addition to the 256 bases of flanking cellular DNA shown below,
            approximately 2500 bases of 3' flanking material were determined.
               Author address R.Desrosiers,
            New England Regional Primate Research Center
            Harvard Medical School
            Southborough MA 01772.
FEATURES             Location/Qualifiers
     source          1. .10535
                     /organism="Simian immunodeficiency virus"
                     /isolate="239"
                     /db_xref="taxon:11723"
     LTR             257. .1074
                     /note="5'LTR"
     LTR             257. .773
                     /note="U3"
     repeat_region   774. .950
                     /note="R"
     misc_feature    775
                     /note="mRNA start (putative); putative"
     LTR             951. .1074
                     /note="U5"
     CDS             1309. .2841
                     /note="gag polyprotein"
                     /codon_start=1
                     /protein_id="AAA47632.1"
                     /db_xref="GI:334650"
                     /translation="MGVRNSVLSGKKADELEKIRLRPNGKKKYMLKHVVWAANELDRF
                     GLAESLLENKEGCQKILSVLAPLVPTGSENLKSLYNTVCVIWCIHAEEKVKHTEEAKQ
                     IVQRHLVVETGTTETMPKTSRPTAPSSGRGGNYPVQQIGGNYVHLPLSPRTLNAWVKL
                     IEEKKFGAEVVPGFQALSEGCTPYDINQMLNCVGDHQAAMQIIRDIINEEAADWDLQH
                     PQPAPQQGQLREPSGSDIAGTTSSVDEQIQWMYRQQNPIPVGNIYRRWIQLGLQKCVR
                     MYNPTNILDVKQGPKEPFQSYVDRFYKSLRAEQTDAAVKNWMTQTLLIQNANPDCKLV
                     LKGLGVNPTLEEMLTACQGVGGPGQKARLMAEALKEALAPVPIPFAAAQQRGPRKPIK
                     CWNCGKEGHSARQCRAPRRQGCWKCGKMDHVMAKCPDRQAGFLGLGPWGKKPRNFPMA
                     QVHQGLMPTAPPEDPAVDLLKNYMQLGKQQREKQRESREKPYKEVTEDLLHLNSLFGG
                     DQ"
     CDS             <2484. .5666
                     /note="pol polyprotein"
                     /codon_start=1
                     /protein_id="AAA47633.2"
                     /db_xref="GI:7583329"
                     /translation="VLELWERGTLCKAMQSPKKTGMLEMWKNGPCYGQMPRQTGGFFR
                     PWSMGKEAPQFPHGSSASGADANCSPRGPSCGSAKELHAVGQAAERKAERKQREALQG
                     GDRGFAAPQFSLWRRPVVTAHIEGQPVEVLLDTGADDSIVTGIELGPHYTPKIVGGIG
                     GFINTKEYKNVEIEVLGKRIKGTIMTGDTPINIFGRNLLTALGMSLNFPIAKVEPVKV
                     ALKPGKDGPKLKQWPLSKEKIVALREICEKMEKDGQLEEAPPTNPYNTPTFAIKKKDK
                     NKWRMLIDFRELNRVTQDFTEVQLGIPHPAGLAKRKRITVLDIGDAYFSIPLDEEFRQ
                     YTAFTLPSVNNAEPGKRYIYKVLPQGWKGSPAIFQYTMRHVLEPFRKANPDVTLVQYM
                     DDILIASDRTDLEHDRVVLQSKELLNSIGFSTPEEKFQKDPPFQWMGYELWPTKWKLQ
                     KIELPQRETWTVNDIQKLVGVLNWAAQIYPGIKTKHLCRLIRGKMTLTEEVQWTEMAE
                     AEYEENKIILSQEQEGCYYQEGKPLEATVIKSQDNQWSYKIHQEDKILKVGKFAKIKN
                     THTNGVRLLAHVIQKIGKEAIVIWGQVPKFHLPVEKDVWEQWWTDYWQVTWIPEWDFI
                     STPPLVRLVFNLVKDPIEGEETYYTDGSCNKQSKEGKAGYITDRGKDKVKVLEQTTNQ
                     QAELEAFLMALTDSGPKANIIVDSQYVMGIITGCPTESESRLVNQIIEEMIKKSEIYV
                     AWVPAHKGIGGNQEIDHLVSQGIRQVLFLEKIEPAQEEHDKYHSNVKELVFKFGLPRI
                     VARQIVDTCDKCHQKGEAIHGQANSDLGTWQMDCTHLEGKIIIVAVHVASGFIEAEVI
                     PQETGRQTALFLLKLAGRWPITHLHTDNGANFASQEVKMVAWWAGIEHTFGVPYNPQS
                     QGVVEAMNHHLKNQIDRIREQANSVETIVLMAVHCMNFKRRGGIGDMTPAERLINMIT
                     TEQEIQFQQSKNSKFKNFRVYYREGRDQLWKGPGELLWKGEGAVILKVGTDIKVVPRR
                     KAKIIKDYGGGKEVDSSSHMEDTGEAREVA"
     misc_feature    5340. .5355
                     /note="polypurine tract"
     CDS             5596. .6240
                     /note="vif protein"
                     /codon_start=1
                     /protein_id="AAA47634.1"
                     /db_xref="GI:334652"
                     /translation="MEEEKRWIAVPTWRIPERLERWHSLIKYLKYKTKDLQKVCYVPH
                     FKVGWAWWTCSRVIFPLQEGSHLEVQGYWHLTPEKGWLSTYAVRITWYSKNFWTDVTP
                     NYADILLHSTYFPCFTAGEVRRAIRGEQLLSCCRFPRAHKYQVPSLQYLALKVVSDVR
                     SQGENPTWKQWRRDNRRGLRMAKQNSRGDKQRGGKPPTKGANFPGLAKVLGILA"
     CDS             6068. .6406
                     /note="vpX protein"
                     /codon_start=1
                     /protein_id="AAA47635.1"
                     /db_xref="GI:334653"
```

```
                /translation="MSDPRERIPPGNSGEETIGEAFEWLNRTVEEINREAVNHLPREL
                IFQVWQRSWEYWHDEQGMSPSYVKYRYLCLIQKALFMHCKKGCRCLGEGHGAGGWRPG
                PPPPPPPGLA"
CDS             6407. .6712
                /note="vpR protein"
                /codon_start=1
                /protein_id="AAA47636.1"
                /db_xref="GI:334654"
                /translation="MEERPPENEGPQREPWDEWVVEVLEELKEEALKHFDPRLLTALG
                NHIYNRHGDTLEGAGELIRILQRALFMHFRGGCIHSRIGQPGGGNPLSAIPPSRSML"
CDS             join(6558. .6853,9062. .9158)
                /note="tat protein"
                /codon_start=1
                /protein_id="AAA47630.1"
                /db_xref="GI:334648"
                /translation="METPLREQENSLESSNERSSCISEADASTPESANLGEEILSQLY
                RPLEACYNTCYCKKCCYHCQFCFLKKGLGICYEQSRKRRRTPKKAKANTSSASNKPIS
                NRTRHCQPEKAKKETVEKAVATAPGLGR"
exon            <6558. .6853
                /note="tat protein,  (first expressed exon)"
                /number=2
CDS             join(6784. .6853,9062. .9315)
                /note="rev protein"
                /codon_start=1
                /protein_id="AAA47631.1"
                /db_xref="GI:334649"
                /translation="MSNHEREEELRKRLRLIHLLHQTNPYPTGPGTANQRRQRKRRWR
                RRWQQLLALADRIYSFPDPPTDTPLDLAIQQLQNLAIESIPDPPTNTPEALCDPTEDS
                RSPQD"
exon            <6784. .6853
                /note="rev protein,  (first expressed exon)"
                /number=2
CDS             6860. .9499
                /note="env polyprotein"
                /codon_start=1
                /protein_id="AAA47637.1"
                /db_xref="GI:334655"
                /translation="MGCLGNQLLIAILLLSVYGIYCTLYVTVFYGVPAWRNATIPLFC
                ATKNRDTWGTTQCLPDNGDYSEVALNVTESFDAWNNTVTEQAIEDVWQLFETSIKPCV
                KLSPLCITMRCNKSETDRWGLTKSITTTASTTSTTASAKVDMVNETSSCIAQDNCTGL
                EQEQMISCKFNMTGLKRDKKKEYNETWYSADLVCEQGNNTGNESRCYMNHCNTSVIQE
                SCDKHYWDAIRFRYCAPPGYALLRCNDTNYSGFMPKCSKVVVSSCTRMMETQTSTWFG
                FNGTRAENRTYIYWHGRDNRTIISLNKYYNLTMKCRRPGNKTVLPVTIMSGLVFHSQP
                INDRPKQAWCWFGGKWKDAIKEVKQTIVKHPRYTGTNNTDKINLTAPGGGDPEVTFMW
                TNCRGEFLYCKMNWFLNWVEDRNTANQKPKEQHKRNYVPCHIRQIINTWHKVGKNVYL
                PPREGDLTCNSTVTSLIANIDWIDGNQTNITMSAEVAELYRLELGDYKLVEITPIGLA
                PTDVKRYTTGGTSRNKRGVFVLGFLGFLATAGSAMGAASLTLTAQSRTLLAGIVQQQQ
                QLLDVVKRQQELLRLTVWGTKNLQTRVTAIEKYLKDQAQLNAWGCAFRQVCHTTVPWP
                NASLTPKWNNETWQEWERKVDFLEENITALLEEAQIQQEKNMYELQKLNSWDVFGNWF
                DLASWIKYIQYGVYIVVGVILLRIVIYIVQMLAKLRQGYRPVFSSPPSYFQQTHIQQD
                PALPTREGKERDGGEGGGNSSWPWQIEYIHFLIRQLIRLLTWLFSNCRTLLSRVYQIL
                QPILQRLSATLQRIREVLRTELTYLQYGWSYFHEAVQAVWRSATETLAGAWGDLWETL
                RRGGRWILAIPRRIRQGLELTLL"
exon            9062. .>9315
                /note="rev protein"
                /number=3
exon            9062. .>9158
                /note="tat protein"
                /number=3
CDS             9333. .10124
                /note="nef protein (in-frame stop at 9612)"
                /pseudo
                /codon_start=1
CDS             9333. .9611
                /note="nef protein (premature termination)"
                /codon_start=1
                /protein_id="AAA47638.1"
                /db_xref="GI:334656"
                /translation="MGGAISMRRSRPSGDLRQRLLRARGETYGRLLGEVEDGYSQSPG
                GLDKGLSSLSCEGQKYNQGQYMNTPWRNPAEEREKLAYRKQNMDDIDE"
misc_feature    9701. .9716
                /note="polypurine tract"
LTR             9718. .10535
                /note="3' LTR"
LTR             9718. .10234
                /note="U3"
repeat_region   10235. .10411
                /note="R"
polyA_signal    10388. .10393
LTR             10412. .10535
                /note="U5"
BASE COUNT      3540 a   1988 c   2607 g   2400 t
ORIGIN
```

Query Match 100.0%; Score 2564; DB 14; Length 10535;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2564; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3105 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 3164

Qy      61 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3165 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 3224

Qy     121 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3225 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 3284

Qy     181 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3285 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 3344

Qy     241 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3345 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 3404

Qy     301 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3405 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 3464

Qy     361 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3465 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 3524

Qy     421 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3525 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 3584

Qy     481 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3585 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 3644

Qy     541 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3645 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 3704

Qy     601 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3705 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 3764

Qy     661 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3765 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 3824

Qy     721 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3825 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 3884

Qy     781 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3885 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 3944

Qy     841 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   3945 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 4004

Qy    901 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4005 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 4064

Qy    961 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4065 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 4124

Qy   1021 AAAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4125 AAAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACA 4184

Qy   1081 CATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATA 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4185 CATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATA 4244

Qy   1141 GTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4245 GTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAG 4304

Qy   1201 TGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCA 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4305 TGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCA 4364

Qy   1261 CCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTAT 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4365 CCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTAT 4424

Qy   1321 TATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGAT 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4425 TATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGAT 4484

Qy   1381 AGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAA 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4485 AGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAA 4544

Qy   1441 GCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAA 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4545 GCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAA 4604

Qy   1501 TATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAA 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4605 TATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAA 4664

Qy   1561 ATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAA 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4665 ATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAA 4724

Qy   1621 GGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTC 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4725 GGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTC 4784

Qy   1681 TTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4785 TTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAA 4844

Qy   1741 GAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGT 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4845 GAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGT 4904

Qy   1801 GATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACT 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4905 GATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACT 4964

Qy   1861 TGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCT 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4965 TGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCT 5024

Qy   1921 AGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTT 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5025 AGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTT 5084
```

```
Qy   1981 CTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAAC 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5085 CTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAAC 5144

Qy   2041 TTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGG 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5145 TTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGG 5204

Qy   2101 GTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAAT 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5205 GTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAAT 5264

Qy   2161 CAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTT 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5265 CAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTT 5324

Qy   2221 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTA 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5325 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTA 5384

Qy   2281 ATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTT 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5385 ATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTT 5444

Qy   2341 AAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAG 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5445 AAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAG 5504

Qy   2401 CTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTA 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5505 CTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTA 5564

Qy   2461 CCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGT 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5565 CCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGT 5624

Qy   2521 TCCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 2564
          |||||||||||||||||||||||||||||||||||||||||||
Db   5625 TCCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 5668
```

```
RESULT 8
AF217181
LOCUS       AF217181               9860 bp    RNA     linear   VRL 01-JAN-2002
DEFINITION  Simian-Human immunodeficiency virus SHIV-C2/1, complete genome.
ACCESSION   AF217181
VERSION     AF217181.1  GI:18026826
KEYWORDS    .
SOURCE      Simian-Human immunodeficiency virus.
  ORGANISM  Simian-Human immunodeficiency virus
            Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
            lentivirus group.
REFERENCE   1  (bases 1 to 9860)
  AUTHORS   Sakai,K., Shinohara,K., Takahashi,E., Ami,Y., Sasaki,Y., Suzaki,Y.,
            Nakasone,T. and Honda,M.
  TITLE     Comparative genetic analysis of simian/human immunodeficiency
            viruses (SHIVs) with pathogenic property
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 9860)
  AUTHORS   Sakai,K., Shinohara,K., Takahashi,E. and Honda,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1999) AIDS Research Center, National Institute of
            Infectious Diseases, Gakuen 4-7-1, Musashimurayama, Tokyo 208-0011,
            Japan
FEATURES             Location/Qualifiers
     source          1. .9860
                     /organism="Simian-Human immunodeficiency virus"
                     /virion
                     /strain="SHIV-C2/1"
                     /specific_host="cynomolgus monkey"
                     /db_xref="taxon:57667"
                     /note="obtained from monkey plasma"
```

misc_feature 1. .5938
/note="derived from SIVmac239"
LTR 1. .301
repeat_region 1. .177
/note="R region"
misc_feature 178. .301
/note="U5 region"
gene 536. .2068
/gene="gag"
CDS 536. .2068
/gene="gag"
/codon_start=1
/product="gag protein"
/protein_id="AAL55638.1"
/db_xref="GI:18026827"
/translation="MGVRNSVLSGKKADELEKIRLRPNGKKKYMLKHVVWAANELDRF
GLAESLLENKEGCQKILSVLAPLVPTGSENLKSLYNTVCVIWCIHAEEKVKHTEEAKQ
IVQRHLVVETGTTETMPKTSRPTAPSSGRGGNYPVQQIGGNYVHLPLSPRTLNAWVKL
IEEKKFGAEVVPGFQALSEGCTPYDINQMLNCVGDHQAAMQIIRDIINEEAADWDLQH
PQPAPQQGQLREPSGSDIAGTTSSVDEQIQWMYRQQNPIPVGNIYRRWIQLGLQKCVR
MYNPTNILDVKQGPKEPFQSYVDRFYKSLRAEQTDAAVKNWMTQTLLIQNANPDCKLV
LKGLGVNPTLEEMLTACQGVGGPGQKARLMAEALKEALAPVPIPFAAAQQRGPRKPIK
CWNCGKEGHSARQCRAPRRQGCWKCGKMDHVMAKCPDRQAGFLGLGPWGKKPRNFPMA
QVHQGLIPTAPPEDPAVDLLKNYMQLGKQQREKQRESREKPYKEVTEDLLHLNSLFGG
DQ"
gene <1711. .4893
/gene="pol"
CDS <1711. .4893
/gene="pol"
/codon_start=1
/product="pol protein"
/protein_id="AAL55639.1"
/db_xref="GI:18026828"
/translation="VLELWERGTLCKAMQSPKKTGMLEMWKNGPCYGQMPRQTGGFFR
PWSMGKEAPQFPHGSSASGADTNCSPRGPSCGSAKELHAVGQAAERKAERKQREALQG
GDRGFAAPQFSLWRRPVVTAHIEGQPVEVLLDTGADDSIVTGIELGPHYTPKIVGGIG
GFINTKEYKNVEIEVLGKRIKGTIMTGDTPINIFGRNLLTALGMSLNFPIAKVEPVKV
ALKPGKDGPKLKQWPLSKEKIVALREICEKMEKDGQLEEAPPTNPYNTPTFAIKKKDK
NKWRMLIDFRELNRVTQDFTEVQLGIPHPAGLAKRKRITVLDIGDAYFSIPLDEEFRQ
YTAFTLPSVNNAEPGKRYIYKVLPQGWKGSPAIFQYTMRHVLEPFRKANPDVTLVQYM
DDILIASDRTDLEHDRVVLQLKELLNSIGFSTPEEKFQKDPPFQWMGYELWPTKWKLQ
KIELPQRETWTVNDIQKLVGVLNWAAQIYPGIKTKHLCRLIRGKMTLTEEVQWTEMAE
AEYEENKIILSQEQEGCYYQEGKPLEATVIKSQDNQWSYKIHQEDKILKVGKFAKIKN
THTNGVRLLAHVIQKIGKEAIVIWGQVPKFHLPVEKDVWEQWWTDYWQVTWIPEWDFI
STPPLVRLVFNLVKDPIEGEETYYTDGSCNKQSKEGKAGYITDRGKDKVKVLEQTTNQ
QAELEAFLMALTDSGPKANIIVDSQYVMGIITGCPTESESRLVNQIIEEMIKKSEIYV
AWVPAHKGIGGNQEIDHLVSQGIRQVLFLERIEPAQEEHDKYHSNVKELVFKFGLPRI
VARQIVDTCDKCHQKGEAIHGQANSDLGTWQMDCTHLEGKIIIVAVHVASGFIEAEVI
PQETGRQTALFLLKLAGRWPITHLHTDNGANFASQEVKMVAWWAGIEHTFGVPYNPQS
QGVVEAMNHHLKNQIDRIREQANSVETIVLMAVHCMNFKRRGGIGDMTPAERLINMIT
TEQEIQFQQSKNSKFKNFRVYYREGRDQLWKGPGELLWKGEGAVILKVGTDIKVVPRR
KAKIIKDYGGGKEVDSSSHMEDTGEAREVA"
gene 4823. .5467
/gene="vif"
CDS 4823. .5467
/gene="vif"
/codon_start=1
/product="vif protein"
/protein_id="AAL55640.1"
/db_xref="GI:18026829"
/translation="MEEEKRWIAVPTWRIPERLERWHSLIKYLKYKTKDLQKVCYVPH
FKVGWAWWTCSRVIFPLQEGSHLEVQGYWHLTPEKGWLSTYAVRITWYSKNFWTDVTP
NYADILLHSTYFPCFTAGEVRRAIRGEQLLSCCRFPRAHKYQVPSLQYLALKVVSDVR
SQGENPTWKQWRRDNRRGLRMAKQNSRGDKQRGGKPPTKGANFPGLAKVLGILA"
gene 5295. .5633
/gene="vpx"
CDS 5295. .5633
/gene="vpx"
/codon_start=1
/product="vpx protein"
/protein_id="AAL55641.1"
/db_xref="GI:18026830"
/translation="MSDPRERIPPGNSGEETIGEAFEWLNRTVEEINREAVNHLPREL
IFQVWQRSWEYWHDEQGMSPSYVKYRYLCLIQKALFMHCKKGCRCLGEGHGAGGWRPG
PPPPPPPGLA"
gene 5634. .5939
/gene="vpr"
CDS 5634. .5939
/gene="vpr"
/codon_start=1
/product="vpr protein"
/protein_id="AAL55642.1"
/db_xref="GI:18026831"
/translation="MEERPPENEGPQREPWDEWVVEVLEELKEEALKHFDPRLLTALG
NHIYNRHGDTLEGAGELIRILQRALFMHFRGGCIHSRIGQPGGGNPLSAIPPSRSML"
misc_feature 5939. .8775
/note="derived from HIV-1"
gene 5948. .8580
/gene="tat"
CDS join(5948. .6162,8487. .8580)
/gene="tat"
/codon_start=1
/product="tat protein"
/protein_id="AAL55643.1"
/db_xref="GI:18026832"
/translation="MEPVDPRLEPWKHPGSKPKTACTNCYCKKCCFHCQVCFTTKALG
ISYGRKKRRQRRRAHQNSQTHQASLSKQPSSQPRGDPTGPKEQKKKVERETETDPVHQ
"
gene 6087. .8761
/gene="rev"
CDS join(6087. .6162,8487. .8761)
/gene="rev"
/codon_start=1
/product="rev protein"
/protein_id="AAL55644.1"
/db_xref="GI:18026833"
/translation="MAGRSGDSDEELIRTVRLIKLLYQSNPPPSLEGTRQARRNRRRR
WRERQRQIRSISERILGTYLGRSAEPVPLQLPPLERLTLDCNEDCGTSGTQGVGSPQI
LVESPTVLESGTKE"
gene 6179. .6424
/gene="vpu"
CDS 6179. .6424
/gene="vpu"
/codon_start=1
/product="vpu protein"
/protein_id="AAL55645.1"
/db_xref="GI:18026834"
/translation="MQPIQIAIVALVVAIIIAIVVWSIVIIEYRKILRQRKIDRLIDR
LIERAEDSGNESEGEISALVEMGVEMGHHAPWDVDDL"
gene 6339. .8948
/gene="env"
CDS 6339. .8948
/gene="env"
/codon_start=1
/product="env polyprotein"
/protein_id="AAL55646.1"
/db_xref="GI:18026835"
/translation="MRVKEKYQHLWRWGWRWGTMLLGMLMICSATEKLWVTVYYGVPV
WREATTTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEVVLGNVTENFNMWKNNMVD
QMHEDIISLWDESLKPCVKLTPLCVTLNCTNLNITKNTTNLTSSSWGMMEEGEIKNCS
FYITTSIRNKVKKEYALFNRLDVVPVKNTSNTKYRLISCNTSVITQACPKVSFQPIPI
HYCVPAGFAILKCNNKTFNGSGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAEEDIVIR
SEDFTDNVKTIIVQLNESVVINCTRPNNNTRERLSIGPGRAFYARRNIIGDIRQAHCN
ISRAKWNNTLQQIVIKLREKFRNKTIAFNQSSGGDPEIVMHSFNCGGEFFYCNTAQLF
NSTWNVAGGTNGTEGNDIITLQCRIKQIINMWQKVGKAMYAPPITGQIRCSSNITGLL
LTRDGGNSTETETEIFRPGGGDMRDNWRSELYKYKVVRIEPIGVAPTRAKRRTVQREK
RAVGIGAVFLGFLGAAGSTMGAASVTLTVQARLLLSGIVQQQNNLLRAIEAQQNMLRL
TVWGIKQLQARVLALERYLRDQQLMGIWGCSGKLICTTSVPWNVSWSNKSVDDIWNNM
TWMEWEREIDNYTDYIYDLLEKSQTQQEKNEKELLELDKWASLWNWFDITNWLWYIRL
FIMIVGGLIGLRIVFAVLSIVNRVRQGYSPLSFQTLLPASRGPDRPEGTEEEGGERDR
DRSGPSVNGSLALIWADLRSLCLFSYHRLRDLLLLIVTRTVELLGRRGWEALKYWWNLL
QYWSQELKNSAVSLLQYGWSYFHEAVQAVWRSATETLAGAWGDVWETLRRGGRWILAI
PRRIRQGLELTLL"
misc_feature 8776. .9860
/note="derived from SIVmac239"
gene 8782. .9573

```
                     /gene="nef"
     CDS             8782. .9573
                     /gene="nef"
                     /codon_start=1
                     /product="nef protein"
                     /protein_id="AAL55647.1"
                     /db_xref="GI:18026836"
                     /translation="MGGAISMRRSRPSGDLRQRLLRARGETYGRLLGEVEDGYSQSPG
                     GLDKGLSSLSCEGQKYNQGQYMNTPWRNPAEEREKLAYRKQNMDDIDEEDDDLVGVSV
                     RPKVPLKTMSYKLAIDMSHFIKEKGGLEGIYYSARRHRILDIYLEKEEGIIPDWQDYT
                     SGPGIRYPKTFGWLWKLVPVNVSDEAQEDEEHYLMHPAQTSQWDDPWGEVLAWKFDPT
                     LAYTYEAYVRYPEEFGSKSGLSEEEVRRRLTARGLLNMADKKETR"
     LTR             9167. .9860
     misc_feature    9167. .9683
                     /note="U3 region"
     repeat_region   9684. .9860
                     /note="R region"
BASE COUNT     3415 a   1789 c   2466 g   2190 t
ORIGIN

  Query Match             99.8%;  Score 2559.2;  DB 14;  Length 9860;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 2561;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2332 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 2391

Qy     61 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2392 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 2451

Qy    121 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2452 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 2511

Qy    181 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2512 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 2571

Qy    241 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2572 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 2631

Qy    301 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2632 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 2691

Qy    361 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2692 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 2751

Qy    421 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2752 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 2811

Qy    481 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2812 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 2871

Qy    541 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2872 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 2931

Qy    601 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 660
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   2932 AGGGTAGTTTTACAGTTAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 2991

Qy    661 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2992 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 3051

Qy    721 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3052 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 3111

Qy     781 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3112 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 3171

Qy     841 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3172 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 3231

Qy     901 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3232 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 3291

Qy     961 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3292 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 3351

Qy    1021 AAAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACA 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3352 AAAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACA 3411

Qy    1081 CATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATA 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3412 CATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATA 3471

Qy    1141 GTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAG 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3472 GTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAG 3531

Qy    1201 TGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCA 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3532 TGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCA 3591

Qy    1261 CCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTAT 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3592 CCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTAT 3651

Qy    1321 TATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGAT 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3652 TATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGAT 3711

Qy    1381 AGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAA 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3712 AGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAA 3771

Qy    1441 GCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAA 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3772 GCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAA 3831

Qy    1501 TATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAA 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3832 TATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAA 3891

Qy    1561 ATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAA 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3892 ATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAA 3951

Qy    1621 GGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTC 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3952 GGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTC 4011

Qy    1681 TTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAA 1740
           |||||||||| |||||||||||||||||| ||||||||||||||||||||||||||||||
Db   4012 TTCTTGGAAAGGATAGAGCCAGCACAAGAGGAACATGATAAATACCATAGTAATGTAAAA 4071

Qy    1741 GAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGT 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4072 GAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGT 4131

Qy    1801 GATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACT 1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    4132 GATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACT 4191

Qy    1861 TGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCT 1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4192 TGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCT 4251

Qy    1921 AGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTT 1980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4252 AGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTT 4311

Qy    1981 CTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAAC 2040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4312 CTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAAC 4371

Qy    2041 TTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGG 2100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4372 TTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGG 4431

Qy    2101 GTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAAT 2160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4432 GTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAAT 4491

Qy    2161 CAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTT 2220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4492 CAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTT 4551

Qy    2221 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTA 2280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4552 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTA 4611

Qy    2281 ATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTT 2340
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4612 ATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTT 4671

Qy    2341 AAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAG 2400
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4672 AAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAG 4731

Qy    2401 CTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTA 2460
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4732 CTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTA 4791

Qy    2461 CCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGT 2520
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4792 CCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGT 4851

Qy    2521 TCCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 2564
           |||||||||||||||||||||||||||||||||||||||||||
Db    4852 TCCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 4895
```

```
RESULT 9
SIVMM32H
LOCUS       SIVMM32H               10277 bp    DNA     linear   VRL 14-APR-2000
DEFINITION  Simian immunodeficiency virus proviral DNA for env, gag, nef, pol,
            rev, tat, vif, vpr, and vpx proteins, complete cds.
ACCESSION   D01065
VERSION     D01065.1  GI:222602
KEYWORDS    env polyprotein; gag polyprotein; nef protein; pol polyprotein; rev
            protein; tat protein; vif protein; vpr protein; vpx protein.
SOURCE      Simian immunodeficiency virus (isolate:SIVmac32H,
            specific_host:Macaca mulatta) proviral DNA, clone:pJ5.
  ORGANISM  Simian immunodeficiency virus
            Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
            lentivirus group.
REFERENCE   1  (sites)
  AUTHORS   Rud,E.W., Yon,J.R., Larder,B.A., Cook,N., Cranage,M.P. and
            Clarke,B.E.
  TITLE     Infectious molecular clones of SIVmac32H: Nef deletion controls
            ability to reisolate virus from rhesus macaques
  JOURNAL   In Vaccines 92 (1992) In press
REFERENCE   2  (bases 1 to 10277)
  AUTHORS   Rud,E.W., Cranage,M., Yon,J., Quirk,J., Ogilvie,L., Cook,N.,
            Webster,S., Dennis,M. and Clarke,B.E.
  TITLE     Molecular and biological characterization of simian
            immunodeficiency virus macaque strain 32H proviral clones
            containing nef size variants
  JOURNAL   J. Gen. Virol. 75 (Pt 3), 529-543 (1994)
  MEDLINE   94172326
   PUBMED   8126450
COMMENT     Submitted (10-JUN-1991) to DDBJ by:
            Erling W Rud
            Molecular Sciences
            Wellcome Research Laboratories
            Langley Court, South Eden Park Road
            Beckenham, Kent BR3 3BS
            U.K.
            Phone:  081-658-2211 x5279
            Fax:    081-663-3532
            SIVmac32H proviral clone pJ5 contains no premature stop codons in
            any of the known SIV open reading frames. The variant pC8 is
            identical to pJ5 except for the changes indicated in the feature
            table.
FEATURES             Location/Qualifiers
     source          1. .10277
                     /organism="Simian immunodeficiency virus"
                     /proviral
                     /isolate="SIVmac32H"
                     /specific_host="Macaca mulatta"
                     /db_xref="taxon:11723"
                     /clone="pJ5"
                     /note="propagated in human T-cell (cell line JM)"
     LTR             1. .818
                     /note="5'LTR"
     LTR             1. .518
                     /note="5'LTR U3 region"
     protein_bind    415. .424
                     /bound_moiety="NFkB"
     protein_bind    429. .438
                     /bound_moiety="Sp1"
     protein_bind    440. .449
                     /bound_moiety="Sp1"
     protein_bind    451. .460
                     /bound_moiety="Sp1"
     protein_bind    462. .471
                     /bound_moiety="Sp1"
     TATA_signal     488. .494
     misc_feature    518. .519
                     /note="put. mRNA start site"
     repeat_unit     519. .694
                     /rpt_type=terminal
     LTR             695. .818
                     /note="5'LTR U5 region"
     primer_bind     822. .849
                     /note="tRNA (Lys) primer binding site"
     gene            1054. .2586
                     /gene="gag"
     CDS             1054. .2586
                     /gene="gag"
                     /codon_start=1
                     /product="gag polyprotein"
                     /protein_id="BAA00862.1"
                     /db_xref="GI:222607"
                     /translation="MGARNSVLSGKKADELEKIRLRPGGKKKYMLKHVVWAANELDRF
                     GLAESLLENKEGCQKILSVLAPLVPTGSENLKSLYNTVCVIWCIHAEEKVKHTEEAKQ
                     IVQRHLVVETGTAETMPKTSRPTAPSSGRGGNYPVQQIGGNYVHLPLSPRTLNAWVKL
                     IEEKKFGAEVVPGFQALSEGCTPYDINQMLNCVGDHQAAMQIIRDIINEEAADWDLQH
                     PQPAPQQGQLREPSGSDIAGTTSSVDEQIQWMYRQQNPIPVGNIYRRWIQLGLQKCVR
                     MYNPTNILDVKQGPKEPFQSYVDRFYKSLRAEQTDAAVKNWMTQTLLIQNANPDCKLV
                     LKGLGVNPTLEEMLTACQGVGGPGQKARLMAEALKEALAPVPIPFAAAQQRGPRKPIK
                     CWNCGKEGHSARQCRAPRRQGCWKCGKMDHVMAKCPDRQAGFLGLGPWGKKPRNFPMA
                     QVHQGLTPTAPPEDPAVDLLKNYMQLGKQQREKQRESREKPYKEVTEDLLHLNSLFGG
                     DQ"
     gene            2229. .5411
                     /gene="pol"
     CDS             <2229. .5411
```

```
                /gene="pol"
                /function="protease, reverse transcriptase, endonuclease"
                /codon_start=1
                /product="pol polyprotein"
                /protein_id="BAA00863.1"
                /db_xref="GI:2160378"
                /translation="VLELWERRTLCKAMQSPKKTGMLEMWKNGPCYGQMPRQTGGFFR
                PWSMGKEAPQFPHGSSASGADANCSPRGPSCGSAKELHAAGQAAERKAERKQREALQG
                GDRGFAAPQFSLWRRPVVTAHIEGQPVEVLLDTGADDSIVTGIELGLHYTPKIVGGIG
                GFINTKEYKNVEIEVLGKRIKGTIMTGDTPINIFGRNLLTALGMSLNLPIAKVEPVKV
                ALKPGKDGPKLKQWPLSKEKIVALREICEKMEKDGQLEEAPPTNPYNTPTFAIKKKDK
                NKWRMLIDFRELNRVTQDFTEVQLGIPHPAGLAKRKRITVLDIGDAYFSIPLDEEFRQ
                YTAFTLPSVNNAEPGKRYIYKVLPQGWKGSPAIFQYTMRHVLEPFRKANPDVTLVQYM
                DDILIASDRTDLEHDRVVLQLKELLNSIGFSTPEEKFQKDPPFQWMGYELWPTKWKLQ
                KIELPQRETWTVNDIQKLVGVLNWAAQIYPGIKTKHLCRLIRGKMTLTEEVQWTEMAE
                AEYEENKIILSQEQEGCYYQEGKPLEATVIKSQDNQWSYKIHQEDKILKVGKFAKIKN
                THTNGVRLLAHVIQKIGKEAIVIWGQVPKFHLPVEKDVWEQWWTDYWQVTWIPEWDFI
                STPPLVRLVFNLVKDPIKGEETYYTDGSCNKQSKEGKAGYITDRGKDKVKVLEQTTNQ
                QAELEAFLMALTDSGPKANIIVDSQYVMGIITGCPTESESRLVNQIIEEMIKKSEIYV
                AWVPAHKGIGGNQEIDHLVSQGIRQVLFLEKIEPAQEEHDKYHSNVKELVFKFGLPRI
                VARQIVDTCDKCHQKGEAIHGQVNSDLGTWQMDCTHLEGKIIIVAVHVASGFIEAEVI
                PQETGRQTALFLLKLAGRWPVTHLHTDNGANFASQEVKMVAWWAGIEHTFGVPYNPQS
                QGVVEAMNHHLKNQIDRIREQANSVETIVLMAVHCMNFKRRGGIGDMTPAERLINMIT
                TEQEIQFQQSKNSKFKNFRVYYREGRDQLWKGPGELLWKGEGAVILKVGTDIKVVPRR
                KAKIIKDYGGGKEVDSSSHMEDTGEAREVA"
repeat_region   2489. .2522
                /rpt_type=direct
                /rpt_unit=2489. .2500
                /rpt_unit=2501. .2512
                /rpt_unit=2513. .1522
gene            5341. .5985
                /gene="vif"
CDS             5341. .5985
                /gene="vif"
                /function="virion infectivity factor"
                /codon_start=1
                /product="vif protein"
                /protein_id="BAA00864.1"
                /db_xref="GI:222609"
                /translation="MEEGKRWIAVPTWRIPERLERWHSLIKYLKYKTKDLQKVCYVPH
                YKVGWAWWTCSRVIFPLQEGSHLEVQGYWHLTPERGWLSTYAVRITWYSRNFWTDVTP
                DYADILLHSTYFPCFTAGEVRRVIRGEQLLSCCKFPRAHRYQVPSLQYLALRVVSYVR
                SQGENPTWKQWRRDSRRSLRMAKQNSRGDKQRGGKPPTKGANFPGLAKVLGILA"
gene            5813. .6151
                /gene="vpx"
CDS             5813. .6151
                /gene="vpx"
                /codon_start=1
                /product="vpx protein"
                /protein_id="BAA00865.1"
                /db_xref="GI:222610"
                /translation="MSDPRERIPPGNSGEETVGEAFEWLNRTVEEINREAVNHLPREL
                IFQVWQRSWEYWHDEQGMSQSYVKYRYLCLIQKALFMHCKKGCRCLGEGHGAGGWRPG
                PPPPPPPGLA"
gene            6152. .6457
                /gene="vpr"
CDS             6152. .6457
                /gene="vpr"
                /codon_start=1
                /product="vpr protein"
                /protein_id="BAA00866.1"
                /db_xref="GI:222611"
                /translation="MEERPPENEGPQREPWDEWVVEVLEELKEEALKHFDPRLLTALG
                NHIYNRHGDTLEGAGELIRILQRALFMHFRGGCNHSRIGQPGGGNPLSTIPPSRGVL"
gene            join(6303. .6598,8801. .8900)
                /gene="tat"
CDS             join(6303. .6598,8801. .8900)
                /gene="tat"
                /function="transactivation"
                /codon_start=1
                /product="tat protein"
                /protein_id="BAA00867.1"
                /db_xref="GI:222603"
                /translation="METPLREQENSLESSNERSSCISEADATTPESANLGEEILSQLY
                RPLEACYNTCYCKKCCYHCQFCFLKKGLGICYEQSRKRRRTPKKAKANTSSASNNRPI
                SNRTRHCQPEKAKKETVEKAVATAPGLGR"
CDS             join(6303. .6598,8804. .8900)
                /gene="tat"
                /note="alternative splicing"
                /codon_start=1
                /product="tat protein"
                /protein_id="BAA00868.1"
                /db_xref="GI:222605"
                /translation="METPLREQENSLESSNERSSCISEADATTPESANLGEEILSQLY
                RPLEACYNTCYCKKCCYHCQFCFLKKGLGICYEQSRKRRRTPKKAKANTSSASNKPIS
                NRTRHCQPEKAKKETVEKAVATAPGLGR"
exon            <6303. .6598
                /gene="tat"
                /product="tat protein"
                /number=1
gene            join(6529. .6598,8801. .9057)
                /gene="rev"
CDS             join(6529. .6598,8801. .9057)
                /gene="rev"
                /codon_start=1
                /product="rev protein"
                /protein_id="BAA00869.1"
                /db_xref="GI:222604"
                /translation="MSSHEREEELRKRLRLIHLLHQTTDPYPTGPGTANQRRQRRRRW
                RRRWQQLLALADRIYSFPDPPTDTPLDLAIQQLQNLAIESIPDPPTNTPEALCGPTEN
                SRSPQD"
CDS             join(6529. .6598,8804. .9057)
                /gene="rev"
                /note="alternative splicing"
                /codon_start=1
                /product="rev protein"
                /protein_id="BAA00870.1"
```

```
 Query Match             98.3%;  Score 2519.8;  DB 14;  Length 10277;
 Best Local Similarity   98.9%;  Pred. No. 0;
 Matches 2536;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy       2 TTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCAA 61
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2851 TTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCAA 2910

Qy      62 AATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAAA 121
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2911 AATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAAA 2970

Qy     122 AGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCCA 181
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2971 AGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCCA 3030

Qy     182 CATTTGCTATAAAGAAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAAC 241
          ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db    3031 CATTTGCTATAAAGAAAAAAGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAAC 3090

Qy     242 TAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGAC 301
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    3091 TAAATAGGGTCACTCAGGACTTTACAGAAGTCCAATTAGGAATACCACACCCTGCAGGAC 3150

Qy     302 TAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCTC 361
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    3151 TAGCAAAAAGGAAAAGGATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCTC 3210

Qy     362 TAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAGC 421
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3211 TAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAGC 3270

Qy     422 CAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATCT 481
          ||||||| ||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    3271 CAGGAAAGCGATACATTTATAAGGTTCTGCCTCAGGGATGGAAAGGGTCACCAGCCATCT 3330

Qy     482 TCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACCT 541
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   3331 TCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACCT 3390

Qy    542 TAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGACA 601
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3391 TAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGACA 3450

Qy    602 GGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAGA 661
          |||||||||||||| |||||||||||||||||||||||||||||||||||||||||| |
Db   3451 GGGTAGTTTTACAGTTAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAAA 3510

Qy    662 AATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGGA 721
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3511 AATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGGA 3570

Qy    722 AGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAGT 781
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3571 AGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAGT 3630

Qy    782 TAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTCT 841
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3631 TAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTCT 3690

Qy    842 GTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCAG 901
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db   3691 GTAGGTTAATTAGAGGAAAAATGACTTTAACAGAGGAAGTTCAGTGGACTGAGATGGCAG 3750

Qy    902 AAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTACC 961
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3751 AAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTACC 3810

Qy    962 AAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTATA 1021
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3811 AAGAGGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTATA 3870

Qy   1022 AAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACAC 1081
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   3871 AAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAGTTTGCAAAGATAAAGAATACAC 3930

Qy   1082 ATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATAG 1141
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   3931 ATACCAATGGAGTTAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATAG 3990

Qy   1142 TGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAGT 1201
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   3991 TGATCTGGGGACAGGTTCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAGT 4050

Qy   1202 GGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCAC 1261
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   4051 GGTGGACAGACTATTGGCAGGTAACCTGGATACCAGAATGGGATTTTATCTCAACACCAC 4110

Qy   1262 CGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTATT 1321
          | ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   4111 CACTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAAAGGGAGAAGAAACCTATT 4170

Qy   1322 ATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGATA 1381
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4171 ATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGATA 4230

Qy   1382 GGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAAG 1441
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4231 GGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAAG 4290

Qy   1442 CATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAAT 1501
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4291 CATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAAT 4350

Qy   1502 ATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAAA 1561
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db   4351 ATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAACCAAA 4410

Qy   1562 TAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAAG 1621
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4411 TAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAAG 4470

Qy   1622 GTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTCT 1681
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4471 GTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTCT 4530

Qy   1682 TCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAAG 1741
          |||||||||||||||||||||||||| |||||||||| ||||||||||||||||||||
Db   4531 TCTTGGAAAAGATAGAGCCAGCACAAGAGGAACATGATAAGTACCATAGTAATGTAAAAG 4590

Qy   1742 AATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGTG 1801
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db   4591 AATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACATGTG 4650

Qy   1802 ATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACTT 1861
          |||||||||||| |||||||| |||||||||||||||| ||||||||||||||||||||
Db   4651 ATAAATGTCATCAAAAAGGAGAAGCTATACATGGGCAGGTAAATTCAGATCTAGGGACTT 4710

Qy   1862 GGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCTA 1921
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4711 GGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCTA 4770

Qy   1922 GTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTTC 1981
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   4771 GTGGATTCATAGAAGCAGAAGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTTC 4830

Qy   1982 TGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAACT 2041
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db   4831 TGTTAAAATTGGCAGGCAGATGGCCTGTTACACATCTACACACAGATAATGGTGCTAACT 4890

Qy   2042 TTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGGG 2101
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4891 TTGCCTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGGG 4950

Qy   2102 TACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAATC 2161
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4951 TACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAATC 5010

Qy   2162 AAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTTC 2221
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5011 AAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTTC 5070

Qy   2222 ATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTAA 2281
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5071 ATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTAA 5130

Qy   2282 TTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTTA 2341
          |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   5131 TTAACATGATCACTACAGAACAAGAAATACAATTTCAACAATCAAAAAACTCAAAATTTA 5190

Qy   2342 AAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAGC 2401
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5191 AAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAGC 5250

Qy   2402 TATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTAC 2461
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5251 TATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTAC 5310

Qy   2462 CCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGTT 2521
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   5311 CCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGGAAAGAGGTGGATAGCAGTT 5370

Qy   2522 CCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 2564
          |||||||||||||||||||||||||||||||||||||||||||
Db   5371 CCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 5413
```

```
RESULT 10
SIVMM251
LOCUS       SIVMM251               10277 bp ss-RNA     linear   VRL 29-NOV-2000
DEFINITION  Simian (macaque) immunodeficiency virus, complete genome; isolate
            Mm251, proviral DNA.
ACCESSION   M19499 M15897 M16125 M24614 Y00294
VERSION     M19499.1  GI:334657
```

KEYWORDS env protein; envelope-associated protein; gag polyprotein; nef protein; pol polyprotein; rev protein; tat protein; vif protein; vpr protein; vpx protein.
SOURCE Simian immunodeficiency virus.
ORGANISM Simian immunodeficiency virus
Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate lentivirus group.
REFERENCE 1 (bases 5731 to 10277)
AUTHORS Hirsch,V.M.
JOURNAL Unpublished
REFERENCE 2 (bases 809 to 1677; 5731 to 10249)
AUTHORS Hahn,B.H., Kumar,P., Taylor,M.E., Arya,S.K. and Shaw,G.M.
TITLE Human retrovirses, cancer and AIDS: approach to prevention and therapy
JOURNAL Unpublished
REFERENCE 3 (bases 6078 to 10249)
AUTHORS Hirsch,V., Riedel,N. and Mullins,J.I.
TITLE The genome organization of STLV-3 is similar to that of the AIDS virus except for a truncated transmembrane protein
JOURNAL Cell 49 (3), 307-319 (1987)
MEDLINE 87187627
PUBMED 3646094
REFERENCE 4 (bases 506 to 5975; 6319 to 10125)
AUTHORS Franchini,G., Gurgo,C., Guo,H.G., Gallo,R.C., Collalti,E., Fargnoli,K.A., Hall,L.F., Wong-Staal,F. and Reitz,M.S. Jr.
TITLE Sequence of simian immunodeficiency virus and its relationship to the human immunodeficiency viruses
JOURNAL Nature 328 (6130), 539-543 (1987)
MEDLINE 87287229
PUBMED 3497350
REFERENCE 5 (bases 7757 to 8090; 8108 to 8225; 8237 to 8625; 8706 to 8898)
AUTHORS Kestler,H.W.
JOURNAL Unpublished
REFERENCE 6 (bases 1 to 10249)
AUTHORS Donahue,P.R., Kornfeld,H., Gallo,M.V. and Mullins,J.I.
JOURNAL Unpublished
REFERENCE 7 (sites)
AUTHORS Kestler,H.W. III, Li,Y., Naidu,Y.M., Butler,C.V., Ochs,M.F., Jaenel,G., King,N.W., Daniel,M.D. and Desrosiers,R.C.
TITLE Comparison of simian immunodeficiency virus isolates
JOURNAL Nature 331 (6157), 619-622 (1988)
MEDLINE 88122665
PUBMED 2893293
COMMENT [7] sites.
Kestler et al. [7], (Nature 331, 619-622 (1988)) present strong evidence that the isolates previously referred to as STLV-III AGM ([4], [3], [1]) and HTLV-IV ([2]) are not authentic, but were derived from cell cultures infected with SIV MAC-251. The reference sequence for this entry is from unpublished (1988), with exception of bases
10250-10277 which are from [1].
the vpr coding region has been annotated to end near where the gene ends in the SIVMM142 sequence. A base change from 'a' at position 6443 in SIVMM142 to 't' at corresponding position 6418 in the SIVMM251 sequence results in the coding sequence ending 4 amino acids earlier than it does in SIVMM142. Also, the variation marked at position 6377 ('aa' in K6W, 'aga' in K78) implies different amino acid sequences for those isolates from that point onward in vpr.
The tat start is annotated in agreement with the site in [2] (also agrees with HIV2ROD) sequence rather than at the site marked in [3] for the K78 sequence.
The start codon for rev is marked at position 6504 in accordance with reference [3] for the K78 sequence. This position also agrees with the rev start in the SIVMM142 sequence. However, [2] annotates the rev start at position 6354. Both of these start sites are in the same frame. The sequence annotation shows two alternate 3' splice junction sites at 8743 and 8785. See reference [3]. Clean copy of sequence unpublished (1988) kindly provided by H.W. Kestler to the
EMBL data library (25-FEB-1988). Computer-readable copy of sequence unpublished (1988) kindly provided by J.Mullins (12-SEP-1988). An internal stop codon is present in the env cds at positions
8782-8784 in unpublished (1988) and 8785-8787 in [4],[3],[1], [2],unpublished (1988).
unpublished (1988) has shown proviral clone BK28 to be biologically active,
causing persistent lymphadenopathy and death in one out of four rhesus monkeys 17 months after infection. BK28 was originally called HTLV-4 (Nature 326, 610-613 (1987)), but it has since been found to correspond to SIVMM251. The K6W sequence [4], which was derived from isolate Mm251, contains about 25 frame shifts in gag-pol relative to this sequence.
A single base insertion at position 9678 results in a frame-shift in the nef coding region. The nef coding region in [4] would produce a protein about 36 residues longer than that produced by the other isolates.
Author [7] address Kestler H.W., Harvard Medical School, New England Regional Primate Research Center, Department of Microbiology, One Pine Hill Drive, Southborough, Mass. 01772, USA.
[7] submitted (25-FEB-1988) to the EMBL data library by. Complete source information:
Simian (macaque) immunodeficiency virus proviral DNA, SIV mac isolate K6W, clones lambda-16, 27, 35, and 11 [4]; and SIV mac isolate K78, clones K2-10 and K2-5 [3], [1]; HTLV4/PK82 and HTLV4/PK190 proviral DNA clones [2]; SIV mac integrated proviral DNA isolate 251, clone lambda 251 unpublished (1988); proviral DNA, isolate Mm251,
clone BK28 unpublished (1988).

FEATURES Location/Qualifiers
source 1. .10277
/organism="Simian immunodeficiency virus"
/proviral
/isolate="Mm251"
/db_xref="taxon:11723"
LTR 1. .806
/note="5' LTR"
misc_binding 428. .437
/note="putative"
/bound_moiety="Sp1 III"
misc_binding 439. .448
/note="putative"
/bound_moiety="Sp1 II"
misc_binding 450. .459
/note="putative"
/bound_moiety="Sp1"
repeat_region 506. .682
/note="R repeat 5' copy"
prim_transcript 507. .10125
/note="genomic mRNA"
prim_transcript 507. .10125
/note="tat, rev, nef subgenomic mRNA"
variation 517. .519
/note="cgg in unpublished (1988); gc in [4]"
/replace="gc"
variation 617. .618
/note="tg in unpublished (1988); t in [4]"
/replace="t"
variation 668
/note="g in unpublished (1988); gc in [4]"
/replace="gc"
variation 694
/note="g in unpublished (1988); c in [4]"
/replace="c"
variation 779
/note="a in unpublished (1988); g in [4]"
/replace="g"
variation 806
/note="a in unpublished (1988); agt in [4]"
/replace="agt"
misc_binding 810. .827
/bound_moiety="Lys-tRNA primer"
variation 825. .826
/note="ac in [4],unpublished (1988); gac in [2]"
/replace="gac"
variation 843. .844

```
                     /note="ga in [2],unpublished (1988); agaa in [4]"
                     /replace="agaa"
     variation       858. .859
                     /note="ct in [2],unpublished (1988); cct in [4]"
                     /replace="cct"
     variation       870
                     /note="a in [4],unpublished (1988); g in [Unpublished
                     (1987) UCLA Symposia on Molecular and Cellular Biology]"
                     /replace="g"
     variation       895. .897
                     /note="gac in [4],unpublished (1988); gc in [2]"
                     /replace="gc"
     variation       919
                     /note="g in [2],unpublished (1988); a in [4]"
                     /replace="a"
     variation       966. .968
                     /note="gtt in [4],unpublished (1988); gt in [2]"
                     /replace="gt"
     variation       1009. .1011
                     /note="atc in [2],unpublished (1988); ac in [4]"
                     /replace="ac"
     variation       1014
                     /note="g in [4],unpublished (1988); a in [2]"
                     /replace="a"
     CDS             1041. .2561
                     /note="gag polyprotein"
                     /codon_start=1
                     /protein_id="AAB59905.1"
                     /db_xref="GI:334660"
                     /translation="MGARNSVLSGKKADELEKIRLRPGGKKKYMLKHVVWAANELDRF
                     GLAESLLENKEGCQKILSVLAPLVPTGSENLKSLYNTVCVIWCIHAEEKVKHTEEAKQ
                     IVQRHLVVETGTAETMPKTSRPTAPSSGRGGNYPVQQIGGNYVHLPLSPRTLNAWVKL
                     IEEKKFGAEVVPGFQALSEGCTPYDINQMLNCVGDHQAAMQIIRDIINEEAADWDLQH
                     PQPAPQQGQLREPSGSDIAGTTSSVDEQIQWMYRQQNPIPVGNIYRRWIQLGLQKCVR
                     MYNPTNILDVKQGPKEPFQSYVDRFYKSLRAEQTDAAVKNWMTQTLLIQNANPDCKLV
                     LKGLGVNPTLEEMLTACQGVGGPGQKARLMAEALKEALAPVPIPFAAAQKRGPRKPIK
                     CWNCGKEGHSARQCRAPRRQGCWKCGKMDHVMAKCPDRQAGFLGLGPWGKKPRNFPMA
                     QVHQGLTPTAPPEDPAVDLLKNYMQLGKQQRESREKPYKEVTEDLLHLNSLFGGDQ"
     variation       1056
                     /note="t in [2],unpublished (1988); g in [4]"
                     /replace="g"
     variation       1083
                     /note="g in [4],unpublished (1988); a in [2]"
                     /replace="a"
     variation       1190
                     /note="g in [2],unpublished (1988); c in [4]"
                     /replace="c"
     variation       1365
                     /note="g in [4],unpublished (1988); a in [2]"

  Query Match             98.1%;  Score 2515;  DB 14;  Length 10277;
  Best Local Similarity   98.8%;  Pred. No. 0;
  Matches 2533;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy       2 TTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCAA 61
          |||||||||||| |||||||||||||||||| |||||||||||||||||| ||||||||
Db    2826 TTCCCATAGCTAAGGTAGAGCCTGTAAAAGTCACCTTAAAGCCAGGAAAGGTTGGACCAA 2885

Qy      62 AATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAAA 121
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2886 AATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAAA 2945

Qy     122 AGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCCA 181
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2946 AGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCCA 3005

Qy     182 CATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAAC 241
          ||||||| |||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    3006 CATTTGCCATAAAGAAAAAAGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAAC 3065

Qy     242 TAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGAC 301
          ||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db    3066 TAAATAGGGTCACTCAGGACTTTACAGAAGTCCAATTAGGAATACCACACCCTGCAGGAC 3125

Qy     302 TAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCTC 361
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    3126 TAGCAAAAAGGAAAAGGATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCTC 3185

Qy     362 TAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAGC 421
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3186 TAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAGC 3245

Qy     422 CAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATCT 481
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3246 CAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATCT 3305

Qy     482 TCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACCT 541
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3306 TCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACCT 3365

Qy     542 TAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGACA 601
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3366 TAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGACA 3425

Qy     602 GGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAGA 661
          |||||||||||||| |||||||||| |||||||||||||| ||||||||||||||||||
Db    3426 GGGTAGTTTTACAGCTAAAGGAACTCTTAAATAGCATAGGGTTCTCTACCCCAGAAGAGA 3485

Qy     662 AATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGGA 721
          ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    3486 AATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCGACAAAATGGA 3545

Qy     722 AGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAGT 781
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3546 AGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAGT 3605

Qy     782 TAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTCT 841
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3606 TAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTCT 3665

Qy     842 GTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCAG 901
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3666 GTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCAG 3725

Qy     902 AAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTACC 961
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    3726 AAGCAGAATATGAGGAAAATAAGATAATTCTCAGTCAGGAACAAGAAGGATGTTATTACC 3785

Qy     962 AAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTATA 1021
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3786 AAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTATA 3845

Qy    1022 AAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACAC 1081
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3846 AAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACAC 3905

Qy    1082 ATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATAG 1141
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    3906 ATACCAATGGAGTTAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATAG 3965

Qy    1142 TGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAGT 1201
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db    3966 TGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAGGGATGTATGGGAACAGT 4025

Qy    1202 GGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCAC 1261
          |||||||||||||||||||||||||||||||||| |||||||||||||||| ||||
Db    4026 GGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAGTGGGATTTTATCTCAACGCCAC 4085

Qy    1262 CGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTATT 1321
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4086 CACTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTATT 4145

Qy    1322 ATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGATA 1381
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4146 ATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGATA 4205
```

```
Qy   1382 GGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAAG 1441
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db   4206 GGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTAGAAG 4265

Qy   1442 CATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAAT 1501
          ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   4266 CATTTCTCATGGCATTGACAGACTCAGGGCCAAAGACAAATATTATAGTAGATTCACAAT 4325

Qy   1502 ATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAAA 1561
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   4326 ATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAACCAAA 4385

Qy   1562 TAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAAG 1621
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4386 TAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAAG 4445

Qy   1622 GTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTCT 1681
          |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db   4446 GTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAGGGGATTAGACAAGTTCTCT 4505

Qy   1682 TCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAAG 1741
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4506 TCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAAG 4565

Qy   1742 AATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGTG 1801
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4566 AATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGTG 4625

Qy   1802 ATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACTT 1861
          ||||||||||||||||||||||| ||||||||||||||| ||||||||||||||||||||
Db   4626 ATAAATGTCATCAGAAAGGAGAAGCTATACATGGGCAGGTAAATTCAGATCTAGGGACTT 4685

Qy   1862 GGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCTA 1921
          |||||||||| |||||||||||||| |||||||| |||||||||||||||||||||||||
Db   4686 GGCAAATGGACTGTACCCATCTAGAAGGAAAAATAGTCATAGTTGCAGTACATGTAGCTA 4745

Qy   1922 GTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTTC 1981
          ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   4746 GTGGATTCATAGAAGCAGAAGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTTC 4805

Qy   1982 TGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAACT 2041
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4806 TGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAACT 4865

Qy   2042 TTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGGG 2101
          |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4866 TTGCCTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGGG 4925

Qy   2102 TACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAATC 2161
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4926 TACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAATC 4985

Qy   2162 AAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTTC 2221
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4986 AAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTTC 5045

Qy   2222 ATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTAA 2281
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5046 ATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTAA 5105

Qy   2282 TTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTTA 2341
          ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db   5106 TTAACATGATCACTACAGAACAAGAAATACAATTTCAACAATCAAAAAACTCAAAATTTA 5165

Qy   2342 AAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAGC 2401
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5166 AAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAGC 5225

Qy   2402 TATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTAC 2461
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5226 TATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTAC 5285

Qy   2462 CCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGTT 2521
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5286 CCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGTT 5345

Qy   2522 CCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 2564
          |||||||||||||||||||||||||||||||||||||||||||
Db   5346 CCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 5388


RESULT 11
SIV1A11AA
LOCUS       SIV1A11AA              10274 bp ss-RNA     linear   VRL 05-MAR-2001
DEFINITION  Simian immunodeficiency virus clone 1A11, complete genome.
ACCESSION   M76764
VERSION     M76764.1  GI:334170
KEYWORDS    .
SOURCE      Simian immunodeficiency virus.
  ORGANISM  Simian immunodeficiency virus
            Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
            lentivirus group.
REFERENCE   1  (bases 1 to 10274)
  AUTHORS   Marthas,M.L., Banapour,B., Sujipto,S., Siegel,M.E., Marx,P.A.,
            Gardner,M.B. and Luciw,P.A.
  TITLE     Rhesus macaques inoculated with molecularly cloned simian
            immunodeficiency virus
  JOURNAL   J. Med. Primatol. 18, 311-319 (1991)
  MEDLINE   89342414
REFERENCE   2  (bases 1 to 10274)
  AUTHORS   Unger,R.E., Stout,M.W. and Luciw,P.A.
  TITLE     Simian immunodeficiency virus (SIVmac) exhibits complex splicing
            for tat, rev, and env mRNA
  JOURNAL   Virology 182 (1), 177-185 (1991)
  MEDLINE   91220650
   PUBMED   2024463
REFERENCE   3  (bases 1 to 10274)
  AUTHORS   Planelles,V., Haigwood,N.L., Marthas,M.L., Mann,K.A., Scandella,C.,
            Lidster,W.D., Shuster,J.R., Van Kuyk,R., Marx,P.A., Gardner,M.B.
            and Luciw,P.A.
  TITLE     Functional and immunological characterization of SIV envelope
            glycoprotein produced in genetically engineered mammalian cells
  JOURNAL   AIDS Res. Hum. Retroviruses 7 (11), 889-898 (1991)
  MEDLINE   92102765
   PUBMED   1760229
REFERENCE   4  (bases 1 to 10274)
  AUTHORS   Luciw,P.A., Shaw,K.E., Unger,R.E., Planelles,V., Stout,M.W.,
            Lackner,J.E., Pratt-Lowe,E., Leung,N.J., Banapour,B. and
            Marthas,M.L.
  TITLE     Genetic and biological comparisons of pathogenic and nonpathogenic
            molecular clones of simian immunodeficiency virus (SIVmac)
  JOURNAL   AIDS Res. Hum. Retroviruses 8 (3), 395-402 (1992)
  MEDLINE   92239216
   PUBMED   1571198
FEATURES             Location/Qualifiers
     source          1..10274
                     /organism="Simian immunodeficiency virus"
                     /proviral
                     /db_xref="taxon:11723"
     LTR             1..817
     TATA_signal     487..492
                     /note="'R = 518'"
     primer_bind     821..848
     CDS             1052..2584
                     /codon_start=1
                     /product="gag polyprotein"
                     /protein_id="AAB54042.1"
                     /db_xref="GI:334173"
                     /translation="MGARNSVLSGKKADELEKIRLRPNGKKKYMLKHVVWAANELDRF
                     GLAESLLENKEGCQKILSVLAPLVPTGSENLKSLYNTVCVIWCIHAEEKVKHTEEAKQ
                     IVQRHLVVETGTAEIMPKTSRPTAPSSGRGGNYPVQQIGGNYVHLPLSPRTLNAWVKL
                     IEEKKFGAEVVPGFQALSEGCTPYDINQMLNCVGDHQAAMQIIRDIINEEAADWDLQH
                     PQPAPQQGQLREPSGSDIAGTTSSVDEQIQWMYRQQNPIPVGNIYRRWIQLGLQKCVR
                     MYNPTNILDVKQGPKEPFQSYVDRFYKSLRAEQTDAAVKNWMTQTLLIQNANPDCKLV
                     LKGLGVNPTLEEMLTACQGVGGPGQKARLMAEALKEALAPVPIPFAAAQQRGPRKPIK
                     CWNCGKEGHSARQCRAPRRQGCWKCGKMDHVMAKCPDRQAGFLGLGPWGKKPRNFPMA
```

```
                QVHQGLTPTAPPEDPAVDLLKNYMQLGKQQREKQRESREKPYKEMTEDLLHLNSLFGG
                DQ"
     CDS        <2227. .5409
                /codon_start=1
                /product="pol polyprotein"
                /protein_id="AAB54043.1"
                /db_xref="GI:2065473"
                /translation="VLELWERGTLCKAMQSPKKTGMLEMWKNGPCYGQMPRQTGGFFR
                PWSMGKEAPQFPHGSSASGADANCSPRGPSCGSAKELHAVGQAAERKTERKQREALQG
                DDRGFAAPQFSLWRRPVVTAHIEGQPVEVLLDTGADDSIVTGIELGPHYTPKIVGGIG
                GFINTKEYKNVEIEVLGKRIRGTIMTGDTPINIFGRNLLTALGMSLNLPIAKVEPVKV
                ALKPGKVGPKLKQWPLSKEKIVALREICEKMEKDGQLEEAPPTNPYNTPTFAIKKKDK
                NKWRMLIDFRELNRVTQDFTEVQLGIPHPAGLAKRKRITVLDIGDAYFSIPLDEEFRQ
                YTAFTLPSVNNAEPGKRYIYKVLPQGWKGSPAIFQYTMRHVLEPFRKANPDVTLVQYM
                DDILIASDRTNLEHDRVVLQLKELLNSIGFSTPEEKFQKDPPFQWMGYELWPTKWKLQ
                KIELPQRETWTVNDIQKLVGVLNWAAQIYPGIKTKHLCRLIRGKMTLTEEVQWTEMAE
                AEYEENKIILSQEQEGCYYQEGKPLEATVIKNQDNQWSYKIHQEDKILKVGKFAKIKN
                THTNGVRLLAHVIQKIGKEAIVIWGQVPKFHLPVEKDVWEQWWTDYWQVTWIPEWNFI
                STPPLVRLVFNLVKDPIEGEETYYTDGSCNKQSKEGKAGYITDRGKDKVKVLEQTTNQ
                QAELEAFLMALTDSGPKANIIVDSQYVMGIITGCPTESESRLVNQIIKEMIKKSEIYV
                AWVPAHKGIGGNQEVDHLVSQGIRQVLFLEKIEPAQEEHDKYHSNVKELVFKFGLPRI
                VARQIVDTCDKCHQKGEAIHGQVNSDLGTWQMDCTHLEGKIIIVAVHVASGFIEAEVI
                PQETGRQTALFLLKLAGRWPITHLHTDNGANFASQEVKMVAWWAGIEHTFGVPYNPQS
                QGVVEAMNHHLKNQIDRIREQANSVETIVLMAVHSMNFKRRGGTGDMTPAERLINMIT
                TEQEIQFQQSKNSKFKNFRVYYREGRDQLWKGPGELLWKGEGAVILKVGTDIKVVPRR
                KAKIIKDYGGGKEMDSSSHMEDTGEAREVA"
     gene       5339. .5983
                /gene="vif"
     CDS        5339. .5983
                /gene="vif"
                /codon_start=1
                /product="vif protein"
                /protein_id="AAB54044.1"
                /db_xref="GI:334175"
                /translation="MEEEKRWIAVPTWRIPERLERWHSLIKYLKYKTKDLQKVCYVPH
                FKVGWAWWTCSRVIFPLQEGSHLEVQGYWHLTPEKGWLSTYAVRVTWYSKNFWTDVTP
                NYADILLHSTYFPCFTAGEVRRAIRGEQLLSCCRFPRAHKYQVPSLQYLALKVVSDVR
                SQGENPTWKQWRRDNRRGLRMAKQNSRGDKQRGGKPPTKGANFPGLAKVLGILA"
     gene       5811. .6149
                /gene="vpr"
     CDS        5811. .6149
                /gene="vpr"
                /codon_start=1
                /product="vpr protein"
                /protein_id="AAB54045.1"
                /db_xref="GI:334176"
                /translation="MSDPRERIPPGNSGEETIGEAFEWLNRTVEEINREAVNHLPREL
                IFQVWQRSWEYWHDEQGMSPSYVKYRYLCLIQKALFMHCKKGCRCLGEGHGAGGWRPG
                PPPPPPPGLA"
     CDS        6150. .6455
                /note="'premature stop @ 6185'"
                /pseudo
                /codon_start=1
     CDS        join(6301. .6596,8799. .8898)
                /codon_start=1
                /product="tat"
                /protein_id="AAB54040.1"
                /db_xref="GI:334171"
                /translation="METPLREQENSLESSSGHSSCTSEAAASTPELANLGEEILSQLY
                RPLEACYNTCYCKKCCYHCQFCFLKKGLGICYEQSRKRRRTPKKAKANTSSASNNRPI
                SNRTRHCQPEKAKKETVEKAVATAPGLGR"
     CDS        join(6527. .6596,8799. .9055)
                /codon_start=1
                /product="rev"
                /protein_id="AAB54041.1"
                /db_xref="GI:334172"
                /translation="MSSHEREEELRKRLRLIHLLHQTIDPYPTGPGTANQRRQRRRRW
                RRRWQQLLALADRIYSFPDPPTDTPLDLAIQQLQNLAIESIPDPPTNTPEALCDPTKD
                SRSPQD"
     CDS        6603. .9239
                /note="env; 'Accession Number M74144(1 to 1572 = 6603 to
                8174; SU)'; 'for splice patterns, see Unger,R.E. et al.';
                'premature stop 1 @ 8801 (SIVmm142)'; 'premature stop 2 @
                9158'"
                /pseudo
                /codon_start=1
     CDS        9073. .9864
                /codon_start=1
                /product="nef"
                /protein_id="AAB54046.1"
                /db_xref="GI:334177"
                /translation="MGGTISMRRSRSTGDLRQRLLRARGETYERLLGEVEDGSSQSLG
                ELDKGLSSLSCEGQKYNQEQYMNTPWRNPAEEREKLAYRKQNMDDIDEEDDDLVGVSV
                RPKVPLRTMSYKLAIDMSHFIKEKGGLEGIYYSARRHRILDIYLEKEEGIIPDWQDYT
                SGPGIRYPKTFGWLWKLVPVNVSDEAQEDEEHYLMHPAQTSQWDDPWGEVPAWKFDPT
                LAYTYEAYVRYPEEFGSKSGLSEEEVRRRLTARGLLNMADKKETR"
     LTR        9458. .10274
     polyA_signal 10127. .10132
                /note="'670. .675 for the 5'LTR'"
BASE COUNT     3475 a   1931 c   2559 g   2309 t
ORIGIN
```

Query Match 97.9%; Score 2510.2; DB 14; Length 10274;
Best Local Similarity 98.7%; Pred. No. 0;
Matches 2530; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

```
Qy      2 TTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCAA 61
          ||||||||||||||||||||||||||||||||||||||||||||||||||| | ||||||||
Db   2849 TTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAAGTTGGACCAA 2908

Qy     62 AATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAAA 121
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2909 AATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAAA 2968

Qy    122 AGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCCA 181
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2969 AGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCCA 3028

Qy    182 CATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAAC 241
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   3029 CATTTGCTATAAAGAAAAAAGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAAC 3088

Qy    242 TAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGAC 301
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   3089 TAAATAGGGTCACTCAGGACTTTACAGAAGTCCAATTAGGAATACCACACCCTGCAGGAC 3148

Qy    302 TAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCTC 361
          |||||||||||||||| |||||||||||||||||||| |||||||||||||||||||||
Db   3149 TAGCAAAAAGGAAAAGGATTACAGTACTGGATATAGGTGACGCATATTTCTCCATACCTC 3208

Qy    362 TAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAGC 421
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3209 TAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAGC 3268

Qy    422 CAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATCT 481
          |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db   3269 CAGGAAAACGATACATTTATAAGGTTCTGCCTCAAGGATGGAAGGGGTCACCAGCCATCT 3328

Qy    482 TCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACCT 541
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3329 TCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACCT 3388

Qy    542 TAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGACA 601
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   3389 TAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAAACCTGGAACATGACA 3448

Qy    602 GGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAGA 661
          |||||||||||||| |||||||||||| ||||||||||||||||||||||||||||||||
Db   3449 GGGTAGTTTTACAGTTAAAGGAACTCTTAAATAGCATAGGGTTTTCTACCCCAGAAGAGA 3508

Qy    662 AATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGGA 721
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3509 AATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGGA 3568

Qy    722 AGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAGT 781
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   3569 AGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAGT 3628

Qy    782 TAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTCT 841
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3629 TAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTCT 3688

Qy    842 GTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCAG 901
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3689 GTAGGCTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCAG 3748

Qy    902 AAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTACC 961
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3749 AAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTACC 3808

Qy    962 AAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTATA 1021
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   3809 AAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAATCAGGACAATCAGTGGTCTTATA 3868

Qy   1022 AAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACAC 1081
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3869 AAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACAC 3928

Qy   1082 ATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATAG 1141
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   3929 ATACCAATGGAGTTAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATAG 3988

Qy   1142 TGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAGT 1201
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3989 TGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAGT 4048

Qy   1202 GGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCAC 1261
          |||||||||||||||||||||||||||||||| |||||| ||||||||||||||||||||
Db   4049 GGTGGACAGACTATTGGCAGGTAACCTGGATACCAGAATGGAATTTTATCTCAACACCAC 4108

Qy   1262 CGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTATT 1321
          | |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4109 CACTAGTGAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTATT 4168

Qy   1322 ATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGATA 1381
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4169 ATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGATA 4228

Qy   1382 GGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAAG 1441
          |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   4229 GGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAGGCAGAATTGGAAG 4288

Qy   1442 CATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAAT 1501
          |||| |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   4289 CATTCCTCATGGCATTAACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAAT 4348

Qy   1502 ATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAAA 1561
          |||||||||||||||||||||||||||||||||||||||||||||| |||||||| ||||
Db   4349 ATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGACTAGTTAACCAAA 4408

Qy   1562 TAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAAG 1621
          ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4409 TAATAAAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAAG 4468

Qy   1622 GTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTCT 1681
          |||||||||||||||||| |||||||||||||||| |||||||||||||||||||||||||
Db   4469 GTATAGGAGGAAACCAAGAAGTAGACCACCTAGTTAGTCAGGGGATTAGACAAGTTCTCT 4528

Qy   1682 TCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAAG 1741
          |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db   4529 TCTTGGAAAAGATAGAGCCAGCACAAGAAGAGCATGATAAATACCATAGTAATGTAAAAG 4588

Qy   1742 AATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGTG 1801
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4589 AATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGTG 4648

Qy   1802 ATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACTT 1861
          |||||||||||||||||||| |||||||||||||||| ||||||||||||||||||||||
Db   4649 ATAAATGTCATCAGAAAGGAGAAGCTATACATGGGCAGGTAAATTCAGATCTAGGGACTT 4708
```

```
Qy   1862 GGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCTA 1921
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4709 GGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCTA 4768

Qy   1922 GTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTTC 1981
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4769 GTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTTC 4828

Qy   1982 TGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAACT 2041
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4829 TGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAACT 4888

Qy   2042 TTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGGG 2101
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4889 TTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGGG 4948

Qy   2102 TACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAATC 2161
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4949 TACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAATC 5008

Qy   2162 AAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTTC 2221
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5009 AAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTTC 5068

Qy   2222 ATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTAA 2281
          |||  ||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   5069 ATTCGATGAATTTTAAAAGAAGGGGAGGAACAGGGGATATGACTCCAGCAGAAAGATTAA 5128

Qy   2282 TTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTTA 2341
          |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db   5129 TTAACATGATCACTACAGAACAAGAAATACAATTTCAACAATCAAAAAACTCAAAATTTA 5188

Qy   2342 AAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAGC 2401
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5189 AAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAGC 5248

Qy   2402 TATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTAC 2461
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5249 TATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTAC 5308

Qy   2462 CCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGTT 2521
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   5309 CCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGATGGATAGCAGTT 5368

Qy   2522 CCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 2564
          |||||||||||||||||||||||||||||||||||||||||||
Db   5369 CCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 5411
```

RESULT 12
RESIVMXX

LOCUS	RESIVMXX 9646 bp DNA linear VRL 01-JUL-1999
DEFINITION	Simian immunodeficiency virus STLV-III(MAC) proviral genome.
ACCESSION	Y00277 M16403
VERSION	Y00277.1 GI:61730
KEYWORDS	art gene; env gene; envelope protein; gag gene; genome; overlapping genes; pol gene; tat gene.
SOURCE	Simian immunodeficiency virus.
ORGANISM	Simian immunodeficiency virus Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate lentivirus group.
REFERENCE	1 (bases 1 to 9646)
AUTHORS	Chakrabarti,L., Guyader,M., Alizon,M., Daniel,M.D., Desrosiers,R.C., Tiollais,P. and Sonigo,P.
TITLE	Sequence of simian immunodeficiency virus from macaque and its relationship to other human and simian retroviruses
JOURNAL	Nature 328 (6130), 543-547 (1987)
MEDLINE	87287230
REFERENCE	2 (bases 1 to 9646)
AUTHORS	Chakrabarti,L.
TITLE	Direct Submission
JOURNAL	Submitted (19-AUG-1987)

```
COMMENT     On Jul 24, 2002 this sequence version replaced gi:334638.
FEATURES             Location/Qualifiers
     source          1. .9646
                     /organism="Simian immunodeficiency virus"
                     /db_xref="taxon:11723"
                     /clone="lambda SIV1"
                     /clone_lib="infected HUT78 cells"
     repeat_region   1. .318
                     /note="5' LTR"
     misc_feature    1. .175
                     /note="LTR (R)"
     misc_feature    176. .318
                     /note="LTR (U5)"
     misc_feature    488. .2068
                     /note="ORF (gag)"
     CDS             551. .2071
                     /note="gag gene product (AA 1-506)"
                     /codon_start=1
                     /protein_id="CAA68379.1"
                     /db_xref="GI:61731"
                     /db_xref="SWISS-PROT:P05894"
                     /translation="MGARNSVLSGKKADELEKIRLRPGGKKKYMLKHVVWAANELDRF
                     GLAESLLENKEGCQKILSVLAPLVPTGSENLKSLYNTVCVIWCIHAEEKVKHTEEAKQ
                     IVQRHLVMETGTAETMPKTSRPTAPFSGRGGNYPVQQIGGNYTHLPLSPRTLNAWVKL
                     IEEKKFGAEVVSGFQALSEGCLPYDINQMLNCVGDHQAAMQIIRDIINEEAADWDLQH
                     PQQAPQQGQLREPSGSDIAGTTSTVEEQIQWMYRQQNPIPVGNIYRRWIQLGLQKCVR
                     MYNPTNILDVKQGPKEPFQSYVDRFYKSLRAEQTDPAVKNWMTQTLLIQNANPDCKLV
                     LKGLGTNPTLEEMLTACQGVGGPGQKARLMAEALKEALAPAPIPFAAAQQKGPRKPIK
                     CWNCGKEGHSARQCRAPRRQGCWKCGKMDHVMAKCPNRQAGFLGLGPWGKKPRNFPMA
                     QVHQGLTPTAPPEEPAVDLLKNYMHLGKQQRESRGKPYKEVTEDLLHLNSLFGGDQ"
     CDS             <1726. .4896
                     /note="pol gene product (AA 1-1056)"
                     /codon_start=1
                     /protein_id="CAA68380.1"
                     /db_xref="GI:1335590"
                     /db_xref="SWISS-PROT:P05896"
                     /translation="VLELWEGRTLCKAMQSPKKTGMLEMWKNGPCYGQMPKQTGGFFR
                     PWPLGKEAPQFPHGSSASGADANCSPRRTSCGSAKELHALGQAAERKQREALQGGDRG
                     FAAPQFSLWRRPVVTAHIEGQPVEVLLDTGADDSIVTGIELGPHYTPKIVGGIGGFIN
                     TKEYKNVEIEVLGKRIKGTIMTGDTPINIFGRNLLTALGMSLNLPIAKVEPVKSPLKP
                     GKDGPKLKQWPLSKEKIVALREICEKMEKDGQLEEAPPTNPYNTPTFAIKKKDKNKWR
                     MLIDFRELNRVTQDFTEVQLGIPHPAGLAKRKRITVLDIGDAYFSIPLDEEFRQYTAF
                     TLPSVNNAEPGKRYIYKVLPQGWKGSPAIFQYTMRHVLEPFRKANPDVTLVQYMDDIL
                     IASDRTDLEHDRVVLQLKELLNSIGFSSPEEKFQKDPPFQWMGYELWPTKWKLQKIEL
                     PQRETWTVNDIQKLVGVLNWAAQIYPGIKTKHLCRLIRGKMTLTEEVQWTEMAEAEYE
                     ENKIILSQEQEGCYYQESKPLEATVIKSQDNQWSYKIHQEDKILKVGKFAKIKNTHTN
                     GVRLLAHVIQKIGKEAIVIWGQVPKFHLPVEKDVWEQWWTDYWQVTWIPEWDFISTPP
                     LVRLVFNLVKDPIEGEETYYVDGSCSKQSKEGKAGYITDRGKDKVKVLEQTTNQQAEL
                     EAFLMALTDSGPKANIIVDSQYVMGIITGCPTESESRLVNQIIEEMIKKTEIYVAWVP
                     AHKGIGGNQEIDHLVSQGIRQVLFLEKIEPAQEEHSKYHSNIKELVFKFGLPRLVAKQ
                     IVDTCDKCHQKGEAIHGQVNSDLGTWQMDCTHLEGKIVIVAVHVASGFIEAEVIPQET
                     GRQTALFLLKLASRWPITHLHTDNGANFASQEVKMVAWWAGIEHTFGVPYNPQSQGVV
                     EAMNHHLKNQIDRIREQANSVETIVLMAVHCMNFKRRGGIGDMTPAERLINMITTEQE
                     IQFQQSKNSKFKNFRVYYREGRDQLWKGPGELLWKGEGAVILKVGTDIKVVPRRKAKI
                     IKDYGGGKEMDSSSHMEDTGEAREVA"
     misc_feature    4793. .5467
                     /note="ORF (Q)"
     CDS             4826. .5470
                     /note="Q gene product (AA 1-214)"
                     /codon_start=1
                     /protein_id="CAA68381.1"
                     /db_xref="GI:61733"
                     /db_xref="SWISS-PROT:P05903"
                     /translation="MEEEKRWIVVPTWRIPERLERWHSLIKYLKYKTKDLQKACYVPH
                     HKVGWAWWTCSRVIFPLQEGSHLEVQGYWNLTPERGWLSTYAVRITWYSKDFWTDVTP
                     EYADILLHSTYFPCFTAGEVRRAIRGERLLSCCRFPRAHKHQVPSLQYLALRVVSHVR
                     SQGENPTWKQWRRDNRRSLRVAKQNSRGDKQRGGKPPTEGANFPGLAKVLGILA"
     misc_feature    5295. .5633
                     /note="ORF (X)"
     CDS             5298. .5636
                     /note="X gene product (AA 1-112)"
                     /codon_start=1
                     /protein_id="CAA68382.1"
                     /db_xref="GI:61734"
                     /db_xref="SWISS-PROT:P05917"
                     /translation="MSDPRERIPPGNSGEETIGEAFEWLNRTVEEINREAVNHLPREL
                     IFQVWQRSWEYWHDEQGMSQSYTKYRYLCLIQKALFMHCKKGCRCLGEGHGAGGWRPG
                     PPPPPPPGLA"
     CDS             5637. .5942
                     /note="R gene product (AA 1-101)"
                     /codon_start=1
                     /protein_id="CAA68383.1"
                     /db_xref="GI:61735"
                     /db_xref="SWISS-PROT:P05958"
                     /translation="MEERPPENEGPQREPWDEWVVEVLKELKEEALKHFDPRLLTALG
                     NHIYNRHGDTLEGAGELIRILQRALFIHFRSGCSHSRIGQPGGGNPLSTIPPSRSML"
     misc_feature    5758. .6083
                     /note="ORF, exon 1 (tat)"
     CDS             join(5788. .6083,8301. .8397)
                     /note="tat gene product"
                     /codon_start=1
                     /protein_id="CAB46521.1"
                     /db_xref="GI:5419939"
                     /translation="METPLREQENSLESSNERSSYISEAAAAIPESANLGEEILSQLY
                     RPLEACYNTCYCKKCCYHCQFCFLKKGLGISYEKSHRRRRTPKKAKANTSSASNEPIP
                     NRIRLCQPKKAKKETVEAAVATAPGLGR"
     misc_feature    6011. .6083
                     /note="ORF, exon 1 (art)"
     CDS             join(6014. .6083,8301. .8554)
                     /note="art gene product"
                     /codon_start=1
                     /protein_id="CAB46522.1"
                     /db_xref="GI:5419940"
                     /translation="MRSHTGEEELRRRLRLIHLLHQTNPYPTGSGSANQRRQKRRRWR
                     QRWQQLLALADRIYSFPDPPTDTPLDLAIQQLQNLAIESIPDPPTNIPEALCDPTENS
                     RSPQA"
     misc_feature    6067. .8735
                     /note="ORF (env)"
     intron          6084. .8300
                     /note="intron I (tat and art) (8301 is 3rd base in codon)"
     mRNA            6090. .8735
                     /note="envelope protein (AA 1-882)"
     misc_feature    8247. .8551
                     /note="ORF, exon 2 (art)"
     misc_feature    8251. .8394
                     /note="ORF, exon 2 (tat)"
     misc_feature    8298. .8300
                     /note="in-frame stop codon (env)"
     misc_feature    8560. .9357
                     /note="ORF (F)"
     CDS             8572. .9360
                     /note="F-gene product (AA 1-262)"
                     /codon_start=1
                     /protein_id="CAA68389.1"
                     /db_xref="GI:61740"
                     /db_xref="SWISS-PROT:P05862"
                     /translation="MGGAISKKRSKPPRDLRQRLLRARGENYGRLFKGVEDGSSQSLG
                     GLDKGLSSLSCEGQKYNQGEYMNTPWRNPAEERKKLPYRKQNIDDIDEEDDDLVGIPV
                     EARVPLRTMSYKLAIDMSHFIKEKGGLEGIYYSARRHRILDIYLEKEEGIIPDWQIHS
                     GPGIRYLKMFGWLWKLIPVNVSDEAQEDEEHYLVHPAQTSQWDDPWGEVLAWKFDPTL
                     AYTYEAYIRYPEEFGSKSGLSEKEVKRRLAARGLLEMADRKETS"
     repeat_region   8957. .9646
                     /note="3' LTR"
     misc_feature    8957. .9471
                     /note="LTR (U3)"
     promoter        9368. .9377
                     /note="enhancer-like sequence"
     promoter        9382. .9391
                     /note="enhancer-like sequence"
     misc_feature    9392. .9402
                     /note="conserved sequence"
     misc_feature    9404. .9412
                     /note="conserved sequence"
     misc_feature    9414. .9424
                     /note="conserved sequence"
     misc_feature    9472. .9646
```

```
                     /note="LTR (R)"
     promoter        9491. .9496
                     /note="pot. promoter sequence"
BASE COUNT     3290 a   1817 c   2387 g   2152 t
ORIGIN

  Query Match             96.3%;  Score 2470.2;  DB 14;  Length 9646;
  Best Local Similarity   97.7%;  Pred. No. 0;
  Matches 2505;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy       2 TTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCAA 61
          |||||||||||| |||||||||||||     |||||||||||||||||||||||||||||
Db    2336 TTCCCATAGCTAAGGTAGAGCCTGTAAAGTCGCCCTTAAAGCCAGGAAAGGATGGACCAA 2395

Qy      62 AATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAAA 121
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2396 AATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAAA 2455

Qy     122 AGATGGAAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCCA 181
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2456 AGATGGAAAAAGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATATAACACCCCCA 2515

Qy     182 CATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAAC 241
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    2516 CATTTGCTATAAAGAAAAAGGATAAAAACAAATGGAGAATGCTGATAGATTTTAGGGAAC 2575

Qy     242 TAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGAC 301
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    2576 TAAATAGGGTCACTCAAGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGAC 2635

Qy     302 TAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCTC 361
          |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db    2636 TAGCAAAAAGGAAAAGGATTACAGTACTGGATATAGGTGACGCATATTTCTCTATACCTC 2695

Qy     362 TAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAGC 421
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2696 TAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAGC 2755

Qy     422 CAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATCT 481
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2756 CAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATCT 2815

Qy     482 TCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACCT 541
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2816 TCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACCT 2875

Qy     542 TAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGACA 601
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2876 TAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGACA 2935

Qy     602 GGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAGA 661
          |||||||||||||| ||| |||||||| ||||||||||||||||  ||||||||||||
Db    2936 GGGTAGTTTTACAGTTAAAAGAACTCTTAAATAGCATAGGGTTTTCATCCCCAGAAGAGA 2995

Qy     662 AATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGGA 721
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    2996 AATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCGACAAAATGGA 3055

Qy     722 AGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAGT 781
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3056 AGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAGT 3115

Qy     782 TAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTCT 841
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3116 TAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTCT 3175

Qy     842 GTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCAG 901
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3176 GTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCAG 3235

Qy     902 AAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTACC 961
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3236 AAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTACC 3295

Qy     962 AAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTATA 1021
          ||||| |||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db    3296 AAGAAAGCAAGCCATTAGAAGCCACGGTGATAAAGAGTCAGGACAATCAGTGGTCTTATA 3355

Qy    1022 AAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACAC 1081
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3356 AAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACAC 3415

Qy    1082 ATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATAG 1141
          |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db    3416 ATACCAATGGAGTTAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATAG 3475

Qy    1142 TGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAGT 1201
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3476 TGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAGT 3535

Qy    1202 GGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCAC 1261
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    3536 GGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTCATCTCAACACCAC 3595

Qy    1262 CGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTATT 1321
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3596 CATTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTATT 3655

Qy    1322 ATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGATA 1381
          ||  |||||||||||||| |||||||||||||||||| |||||||||||||||||||| |
Db    3656 ATGTAGATGGATCATGTAGTAAACAGTCAAAAGAAGGAAAAGCAGGATATATCACAGACA 3715

Qy    1382 GGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAAG 1441
          |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db    3716 GGGGCAAAGACAAGGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAAG 3775

Qy    1442 CATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAAT 1501
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3776 CATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGACTCACAAT 3835

Qy    1502 ATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAAA 1561
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db    3836 ATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAACCAAA 3895

Qy    1562 TAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAAG 1621
          |||||||||||||||| |||||| |||||||||||||| ||||||||||||||||||||||
Db    3896 TAATAGAAGAAATGATCAAAAAGACAGAAATTTATGTGGCATGGGTACCAGCACACAAAG 3955

Qy    1622 GTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTCT 1681
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3956 GTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTCT 4015

Qy    1682 TCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAAG 1741
          |||||||||||||||||||||||||||||||||||  ||||||||||||||  ||||||
Db    4016 TCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATAGTAAATACCATAGTAACATAAAAG 4075

Qy    1742 AATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGTG 1801
          |||||||||||||||||||||||||||||| |||||||| ||||||||||||||  ||||
Db    4076 AATTGGTATTCAAATTTGGATTACCCAGACTAGTGGCCAAACAGATAGTAGACACATGTG 4135

Qy    1802 ATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACTT 1861
          |||||||||||| |||||||| |||||||||||||||| |||||||| ||||||||||||
Db    4136 ATAAATGTCATCAAAAAGGAGAAGCTATACATGGGCAGGTAAATTCAGACCTAGGGACTT 4195

Qy    1862 GGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCTA 1921
          |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db    4196 GGCAAATGGATTGTACCCATCTAGAGGGAAAAATAGTCATAGTTGCAGTACATGTAGCTA 4255

Qy    1922 GTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTTC 1981
          |||||||||||||||||| |||||||||||||| |||||||||||||||||||||||||||
Db    4256 GTGGATTCATAGAAGCAGAAGTAATTCCACAAGAAACAGGAAGACAGACAGCACTATTTC 4315

Qy    1982 TGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAACT 2041
          |||||||||||||| |||||||||||||||||||| |||||||||||||||||||||||||
Db    4316 TGTTAAAATTGGCAAGCAGATGGCCTATTACACATCTGCACACAGATAATGGTGCTAACT 4375
```

```
Qy    2042 TTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGGG 2101
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4376 TTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGGG 4435

Qy    2102 TACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAATC 2161
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4436 TACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAATC 4495

Qy    2162 AAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTTC 2221
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4496 AAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTTC 4555

Qy    2222 ATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTAA 2281
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4556 ATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTAA 4615

Qy    2282 TTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTTA 2341
           |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    4616 TTAACATGATCACTACAGAACAAGAAATACAATTTCAACAATCAAAAAACTCAAAATTTA 4675

Qy    2342 AAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAGC 2401
           ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    4676 AAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAGCTGTGGAAGGGACCCGGTGAGC 4735

Qy    2402 TATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTAC 2461
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    4736 TATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGAACAGACATTAAGGTAGTAC 4795

Qy    2462 CCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGTT 2521
           |||| ||||||||||||||||||||||||||||||||||||||| ||||||| ||||
Db    4796 CCAGGAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGATGGATAGTAGTT 4855

Qy    2522 CCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 2564
           |||||||||||||||||||||||||||||||||||||||||||
Db    4856 CCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 4898
```

RESULT 13
SIU79412

```
LOCUS       SIU79412                9664 bp    DNA     linear   VRL 21-JAN-1998
DEFINITION  Simian immunodeficiency virus isolate SIVMne027, complete genome.
ACCESSION   U79412
VERSION     U79412.1  GI:2737927
KEYWORDS    .
SOURCE      Simian immunodeficiency virus.
  ORGANISM  Simian immunodeficiency virus
            Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
            lentivirus group.
REFERENCE   1  (bases 1 to 9664)
  AUTHORS   Kimata,J.T., Mozaffarian,A. and Overbaugh,J.
  TITLE     A lymph node-derived cytopathic simian immunodeficiency virus Mne
            variant replicates in nonstimulated peripheral blood mononuclear
            cells
  JOURNAL   J. Virol. 72 (1), 245-256 (1998)
  MEDLINE   98080410
   PUBMED   9420221
REFERENCE   2  (bases 1 to 9664)
  AUTHORS   Kimata,J.T., Mozaffarian,A. and Overbaugh,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-1996) Microbiology, University of Washington,
            Health Sciences Building, HSB 320, Seattle, WA 98195-7242, USA
FEATURES             Location/Qualifiers
     source          1. .9664
                     /organism="Simian immunodeficiency virus"
                     /proviral
                     /strain="Mne"
                     /isolate="SIVMne027"
                     /db_xref="taxon:11723"
     LTR             <1. .300
                     /note="5' long terminal repeat"
     repeat_region   1. .175
                     /note="R region"
                     /rpt_type=direct
     repeat_region   176. .300
                     /note="U5 region"
                     /rpt_type=direct
     gene            533. .2053
                     /gene="gag"
     CDS             533. .2053
                     /gene="gag"
                     /codon_start=1
                     /product="gag polyprotein"
                     /protein_id="AAB97129.1"
                     /db_xref="GI:2737928"
                     /translation="MGARNSVLSGKKADELEKIRLRPGGKKKYMLKHVVWAANELDRF
                     GLADSLLENKEGCQKILSALAPLVPTGSENLKSLYNTVCVIWCIHAEEKVKHTEEAKQ
                     IVQRHLVVETGTAETMPKTSRPTAPSSGKRGNYPVQQIGGNYTHLPLSPRTLNAWVKL
                     IEEKKFGAEVVPGFQALSEGCIPYDINQMLNCVGEHQAAMQIIREIINEEAADWDLQH
                     PQQAPQQGQLREPSGSDIAGTTSTVDEQIQWMYRQQNPIPVGNIYRRWIQLGLQKCVK
                     MYNPTNILDIRQGPKEPFQSYVDRFYKSLRAEQTDPAVKNWMTQTLLIQNANPDCKLV
                     LKGLGMNPTLEEMLTACQGVGGPGQKAKLMAEALKEALAPGPLPFAAAQQKGPRKPIK
                     CWNCGKEGHSARQCRAPRRQGCWKCGQMGHVMAKCPDRQAGFLGFGPWGKKPRNFPMA
                     QMHQGLTPTAPPEDPAVDLLKNYMQLGKQQRESKRKPYKEVAEDLLHLNSLFGEDQ"
     gene            1918. .4878
                     /gene="pol"
     CDS             <1918. .4878
                     /gene="pol"
                     /codon_start=1
                     /product="pol polyprotein"
                     /protein_id="AAB97130.1"
                     /db_xref="GI:2737929"
                     /translation="RGPSCGSAKKLHAVGQTAERKQKEALQGGGRGFAAPQFSLWRRP
                     VVTAHIEGQPAEVLLDTGADDSIVAGIELGPHYTPKIVGGIGGFINTKEYKNVEIEVL
                     GKRIKGTIMTGDTPINIFGRNLLTALGMSLNFPIAKVEPVKVTLKPGKDGPKLRQWPL
                     SKEKIVALREICEKMEKDGQLEEAPPTNPYNTPTFAIKKKDKNKWRMLIDFRELNKVT
                     QDFTEVQLGIPHPAGLAKRKRITVLDVGDAYFSIPLDEEFRQYTAFTLPSVNNAEPGK
                     RYIYKVLPQGWKGSPAIFQHTMRNVLEPFRKANPDVTLVQYMDDILVASDRTDLEHDR
                     VVLQLKELLNSIGFSTPEEKFQKDPPFQWMGYELWPTKWKLQKIELPQKETWTVNDIQ
                     KLVGVLNWAAQIYPGIKTKHLCRLIRGKMTLTEEVQWTEMAEAEYEENKIILSQEQEG
                     CYYQEGKPLEATVIKNQDNQWSYKIHQEDKILKVGKFAKIKNTHTNGVRLLAHVIQKI
                     GKEAIVIWGQVPKFHLPVEKDVWEQWWTDYWQVTWIPEWDFISTPPLVRLVFNLVKDP
                     IEGEETYYVDGSCNKQSKEGKAGYITDRGKDKVRVLEQTTNQQAELEAFLMALTDSGP
                     KANIIVDSQYVMGIITGCPTESESRLVNQIIEEMIKKTEIYVAWVPAHKGIGGNQEID
                     HLVSQGIRQVLFLEKIEPAQEEHDKYHSNVKELVFKFGLPRLVAKQIVDTCDKCHQKG
                     EAIHGQVNSDLGTWQMDCTHLEGKIIIVAVHVASGFIEAEVIPQETGRQTALFLLKLA
                     SRWPITHLHTDNGANFTSQEVKMVAWWAGIEHTFGVPYNPQSQGVVEAMNHHLKNQID
                     RIREQANSMETIVLMAVHCMNFKRRGGIGDMTPAERLLNMITTEQEIQFQQSKNSKFK
                     NFRVYYREGRDQLWKGPGELLWKGEGAVILKVGTDIKVVPRRKAKIIKDYGGGKEVDS
                     SSHMEDTGEAREVA"
     gene            4808. .5452
                     /gene="vif"
     CDS             4808. .5452
                     /gene="vif"
                     /codon_start=1
                     /product="vif protein"
                     /protein_id="AAB97131.1"
                     /db_xref="GI:2737930"
                     /translation="MEEEKRWIAVPTWRIPERLERWHSLIKYLKYKTKDLQKVCYVPH
                     HKVGWAWWTCSRVIFPLQEESQLEVQGYWNLTPERGWLSTYAVRITWYSRNFWTDVTP
                     DYADILLHSTYFPCFTAGEVRRAIRGEQLLSCCRFPRAHKNQVPSLQYLALRVVSYVR
                     SQRENPTWKQWRRDNRRSLRMAKQNSRGDKQRGSKPPTKGADFPGLAKVLGILA"
     gene            5280. .5618
                     /gene="vpx"
     CDS             5280. .5618
                     /gene="vpx"
                     /codon_start=1
                     /product="vpx protein"
                     /protein_id="AAB97132.1"
                     /db_xref="GI:2737931"
                     /translation="MSDPRERIPPGNSGEETIGEAFEWLNRTVEEINREAVNHLPREL
                     IFQVWQRSWEYWHDEQGMSPSYVKYRYLCLIQKALFMHCKKGCRCLGEGHGAGGWRPG
                     PPPPPPPGLA"
     gene            5619. .5924
                     /gene="vpr"
     CDS             5619. .5924
                     /gene="vpr"
```

```
                     /codon_start=1
                     /product="vpr protein"
                     /protein_id="AAB97133.1"
                     /db_xref="GI:2737932"
                     /translation="MEERPPEDEGPQREPWDEWVVEVLEELKEEALKHFDPRLLTALG
                     NYIYDRHGDTLEGAGELIKILQRALFMHFRGGCNHSRIGQSGGGNPLSTIPPSRSML"
gene                 5770. .8379
                     /gene="tat"
CDS                  join(5770. .6065,8283. .8379)
                     /gene="tat"
                     /codon_start=1
                     /product="tat protein"
                     /protein_id="AAB97134.1"
                     /db_xref="GI:2737933"
                     /translation="METPLREQENSLKSSNGRSSCTSEAAATTLESANLEEEILSQLY
                     RPLEACYNTCYCKKCCYHCQFCFLKKGLGICYEQSRRRRRTPKKVKANTSSASNKPIS
                     DRTKHCQPKKEQKETVETAVATVPGLGR"
gene                 5996. .8536
                     /gene="rev"
CDS                  join(5996. .6065,8283. .8536)
                     /gene="rev"
                     /codon_start=1
                     /product="rev protein"
                     /protein_id="AAB97135.1"
                     /db_xref="GI:2737934"
                     /translation="MSSHAEEEELRRRLRLIHLLHQTNPYPTGPSTANQRRNRRRRWR
                     QRWQQFLALADRVYSFPDPPTDTPLDLAIQQLQNLAIESIPDPPTNIPEILHDPTESP
                     RSPQD"
gene                 6072. .8720
                     /gene="env"
CDS                  6072. .8720
                     /gene="env"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAB97136.1"
                     /db_xref="GI:2737935"
                     /translation="MGCLGNQLLIAILFLSAYGIYCIQYVTVFYGVPAWRNATIPLFC
                     VTRNRDTWGTTQCLPDNDDYSELALNITESFDAWENTVTEQAIEDVWHLFETSIKPCV
                     KLTPLCITMKCNKSETDKWGLTKSSTTTAAPTTKTTTTKEIEVVNENSTCVNRDNCTG
                     LEQEPMISCKFNMTGLKRDKKREYNETWYSADLVCEQGNSTEDESRCYMNHCNTSVIQ
                     ESCDKHYWDAIRFRYCAPPGYALLRCNDTKYSGFMPNCSKVVVSSCTRMMETQTSTWF
                     GFNGTRAENRTYIYWHSKDNRTIISLNKYNNLTMKCRRPGNKTVLPVTIMSGLVFHSQ
                     PINERPKQAWCRFEGNWKEAIKEVKQTIVKHPRYTGTNNTDKINLTAPRGGDPEVTFM
                     WTNCRGEFLYCKMNWFLNWVEDKNLTGTTQKPQEQHKRNYVPCHIRQIINTWHKVGKN
                     VYLPPREGDLTCNSTVTSLIANIDWIDGNQTNITMSAEVAELYRLELGDYKLVEITPI
                     GLAPTNVKRYTTGGTPRNKRGVFVLGFLGFLATAGSAMGAASLTLTAQSRTLLAGIVQ
                     QQQQLLDVVKRQQELLRLTVWGTKNLQTRVTAIEKYLKDQAQLNAWGCAFRQVCHTTV
                     PWPNASLTPNWNNETWQEWERKVDFLEENITALLEEAQIQQEKNMYELQKLNSWDVFG
                     NWFDLASWIRYIQYGVYIVVGVILLRIVIYIVQMLAKLRQGYRPVFSSPPSYFQQTHI
                     RQDQALPTKEGTEGDGGDSGGNSSWPWQIEYIHFLIRQLIRLLTWLFSNCRTLLSRAY
                     QILQPIFQRFSTTLQRVREVLRTELTYLQYGWSYFQEAVQVAWRSATETLAGAWGDLW
                     ETLGRVGRWILAIPRRIRQELELTLL"
gene                 8554. .9345
                     /gene="nef"
CDS                  8554. .9345
                     /gene="nef"
                     /codon_start=1
                     /product="nef protein"
                     /protein_id="AAB97137.1"
                     /db_xref="GI:2737936"
                     /translation="MGGATSKKRSKSPGDLRQRLLRARGETYGRLWEGLEDGYSQSLG
                     GSDKSLSLLSCEGQKYNQGQFMNTPWKNPAGEREKLAYRKQNMDDIDEEDDGLVGVPV
                     RPRVPLRAMTYKLAVDMSHFIKEKGGLEGIYYSERRHKILDIYLEKEEGIVPDWQDYT
                     SGPGPRYPKTFGWLWKLVPVNVSDEAQEGEENYLLHPAQTSQWDDPWGEVLVWKFDPT
                     LAYTYEAYIRYPEEFGSKSGLSEEEVRRRLTARGLLKMADKKETS"
LTR                  8939. .>9664
                     /note="3' long terminal repeat"
repeat_region        8939. .9457
                     /note="U3 region"
                     /rpt_type=direct
repeat_region        9457. .9632
                     /note="R region"
                     /rpt_type=direct
repeat_region        9633. .9664
                     /note="U5 region"
                     /rpt_type=direct
BASE COUNT     3301 a   1829 c   2387 g   2147 t
ORIGIN

  Query Match             95.8%;  Score 2455.2;  DB 14;  Length 9664;
  Best Local Similarity   97.3%;  Pred. No. 0;
  Matches 2496;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy       1 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 60
             || |||||||||||| |||||||||||||||||| |||||||||||||||| |||||||||
Db    2317 TTCCCCATAGCTAAGGTAGAGCCTGTAAAAGTCACCTTAAAGCCAGGAAAAGATGGACCA 2376

Qy      61 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 120
           ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2377 AAATTGAGGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 2436

Qy     121 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2437 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 2496

Qy     181 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 240
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    2497 ACATTTGCTATAAAGAAAAAGGACAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 2556

Qy     241 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 300
           ||||||| |||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    2557 CTAAATAAGGTCACTCAGGACTTTACAGAAGTCCAATTAGGAATACCACACCCTGCAGGA 2616

Qy     301 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 360
           |||||||||||||| || || |||||||||||| ||||||| ||||||||||||||||||||
Db    2617 CTAGCAAAAAGGAAGAGGATCACAGTACTGGATGTAGGTGACGCATATTTCTCCATACCT 2676

Qy     361 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 420
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db    2677 CTAGATGAAGAATTTAGGCAGTACACTGCTTTTACTTTACCATCAGTAAATAATGCAGAA 2736

Qy     421 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 480
           |||||||||||||||||||||||||| |||||||||| ||||||||||||||||||||||
Db    2737 CCAGGAAAACGATACATTTATAAGGTCCTGCCTCAGGGGTGGAAGGGGTCACCAGCCATC 2796

Qy     481 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 540
           |||||| |||||||||| |||||| ||||||||||||||||||||||||||||||||||||
Db    2797 TTCCAACACACTATGAGAAATGTGCTGGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 2856

Qy     541 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 600
           |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    2857 CTAGTCCAGTATATGGATGACATCTTAGTAGCTAGTGACAGGACAGACCTGGAACATGAC 2916

Qy     601 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 660
           |||||||||||||| |||||||||| |||||||||||||||||||||||||||||||||||
Db    2917 AGGGTAGTTTTACAGTTAAAGGAACTCTTAAATAGCATAGGGTTTTCTACCCCAGAAGAG 2976

Qy     661 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 720
           || |||||||||||||||||||||||||||||||| ||||| |||||||||||||||||||
Db    2977 AAGTTCCAAAAAGATCCCCCATTTCAATGGATGGGGTATGAATTATGGCCAACAAAATGG 3036

Qy     721 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 780
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    3037 AAGTTGCAAAAGATAGAGTTGCCACAAAAAGAGACCTGGACAGTGAATGATATACAGAAG 3096

Qy     781 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3097 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 3156

Qy     841 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3157 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 3216

Qy     901 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 960
           || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3217 GAGGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 3276
```

```
Qy     961 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 1020
           ||||||||||||||||||| |||||||||||||||| |||||||||||||||||||||||
Db    3277 CAAGAAGGCAAGCCATTAGAGGCCACGGTAATAAAGAATCAGGACAATCAGTGGTCTTAT 3336

Qy    1021 AAAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACA 1080
           || ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db    3337 AAGATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAAAATACA 3396

Qy    1081 CATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATA 1140
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    3397 CATACCAATGGAGTTAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATA 3456

Qy    1141 GTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAG 1200
           |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    3457 GTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTCGAGAAGGATGTATGGGAACAG 3516

Qy    1201 TGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCA 1260
           |||||||||||||||||||||||||||||||| |||||||||||||||||||| ||
Db    3517 TGGTGGACAGACTATTGGCAGGTAACCTGGATACCAGAATGGGATTTTATCTCAACGCCG 3576

Qy    1261 CCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTAT 1320
           || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3577 CCACTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTAT 3636

Qy    1321 TATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGAT 1380
           |||  |||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    3637 TATGTAGATGGGTCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGAT 3696

Qy    1381 AGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAA 1440
           |||||||||||||||| | ||| ||||||||||||||||||||||||||||||||||||||
Db    3697 AGGGGCAAAGACAAAGTGAGAGTCTTAGAACAGACTACTAATCAACAAGCAGAATTGGAA 3756

Qy    1441 GCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAA 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3757 GCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAA 3816

Qy    1501 TATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAA 1560
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    3817 TATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAACCAA 3876

Qy    1561 ATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAA 1620
           |||||||||||||||||||||| |||||||||||||||||||||| ||||||||||||
Db    3877 ATAATAGAAGAAATGATTAAAAAGACAGAAATTTATGTAGCATGGGTGCCAGCACACAAA 3936

Qy    1621 GGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTC 1680
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    3937 GGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTCCTC 3996

Qy    1681 TTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAA 1740
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3997 TTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAA 4056

Qy    1741 GAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGT 1800
           |||||||||||||||||||||||||||||| |||||||| |||||||||||||| |||
Db    4057 GAATTGGTATTCAAATTTGGATTACCCAGACTAGTGGCCAAACAGATAGTAGACACATGT 4116

Qy    1801 GATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACT 1860
           |||||||||||||||||||||| |||||||||||||||| ||||||||||||||||||||
Db    4117 GATAAATGTCATCAGAAAGGAGAAGCTATACATGGGCAGGTAAATTCAGATCTAGGGACT 4176

Qy    1861 TGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCT 1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4177 TGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCT 4236

Qy    1921 AGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTT 1980
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    4237 AGTGGATTCATAGAAGCAGAAGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTT 4296

Qy    1981 CTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAAC 2040
           |||||||||||||| |||||||||||||||| |||||||||||||||||||||| |||
Db    4297 CTGTTAAAATTGGCAAGCAGATGGCCTATTACGCATCTACACACAGATAATGGTGCCAAC 4356

Qy    2041 TTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGG 2100
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4357 TTTACTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGG 4416

Qy    2101 GTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAAT 2160
           |||||||||||||||||||||||||||||||||||||||||||||||||| || ||||||
Db    4417 GTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCATCTAAAAAAT 4476

Qy    2161 CAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTT 2220
           |||||||||||||||||||||||||||||||||| | ||||||||||||||||||||||
Db    4477 CAAATAGATAGAATCAGGGAACAAGCAAATTCAATGGAAACCATAGTATTAATGGCAGTT 4536

Qy    2221 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTA 2280
           |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    4537 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGGTTA 4596

Qy    2281 ATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTT 2340
           |||||||||||||||||||||||| |||||||| ||||||||||||||||||||||||||
Db    4597 CTTAACATGATCACTACAGAACAAGAAATACAATTCCAACAATCAAAAAACTCAAAATTT 4656

Qy    2341 AAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAG 2400
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4657 AAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAG 4716

Qy    2401 CTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTA 2460
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4717 CTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTA 4776

Qy    2461 CCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGT 2520
           |||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db    4777 CCCAGAAGAAAGGCTAAGATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGT 4836

Qy    2521 TCCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 2564
           |||||||||||||||||||||||||||||||||||||||||||||
Db    4837 TCCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 4880
```

```
RESULT 14
SIVMNE
LOCUS       SIVMNE                  9628 bp ss-RNA     linear   VRL 05-MAR-1996
DEFINITION  Simian immunodeficiency virus (SIV) complete proviral genome.
ACCESSION   M32741
VERSION     M32741.1  GI:334692
KEYWORDS    complete genome.
SOURCE      Simian immunodeficiency virus (clone:  8.) cDNA to genomic RNA.
  ORGANISM  Simian immunodeficiency virus
            Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
            lentivirus group.
REFERENCE   1  (bases 1 to 9628)
  AUTHORS   Benveniste,R.E., Heidecker,G., Greenwood,J. and Gonda,M.A.
  JOURNAL   Unpublished (1990)
COMMENT     Kindly submitted in computer-readable form by R. Benveniste.  The
            gag protein sequence was reported in J. Virol. 62, 2587-2595, 1988.
            This molecular clone, after transfection into T-cell lines,
            produces infectious viral particles. In particular, clone 8 has
            been inoculated intravenously into two pig-tailed macaques causing
            CD4 lymphocyte depletion; see J. Virol. 62, 2091-2101, 1988.  The
            env cds is truncated as is the case with MM251 and MM142.
              Author address:R.Benveniste
              National Cancer Institute
              Frederick, MD
              (301-698-5836).
FEATURES             Location/Qualifiers
     source          1. .9628
                     /organism="Simian immunodeficiency virus"
                     /proviral
                     /specific_host="Macaca nemestrina"
                     /db_xref="taxon:11723"
                     /clone="8."
     CDS             533. .2053
                     /note="gag polyprotein"
                     /codon_start=1
                     /protein_id="AAA91930.1"
```

```
                     /db_xref="GI:334696"
                     /translation="MGARNSVLSGKKADELEKIRLRPGGKKKYMLKHVVWAANELDRF
                     GLAESLLENKEGCQKILSVLAPLVPTGSENLKSLYNTVCVIWCIHAEEKVKHTEEAKQ
                     IVQRHLVVETGTAETMPKTSRPTAPSSGRGGNYPVQQVGGNYTHLPLSPRTLNAWVKL
                     IEEKKFGAEVVPGFQALSEGCTPYDINQMLNCVGEHQAAMQIIREIINEEAADWDLQH
                     PQQAPQQGQLREPSGSDIAGTTSTVDEQIQWMYRQQNPIPVGNIYRRWIQLGLQKCVR
                     MYNPTSILDVKQGPKEPFQSYVDRFYKSLRAEQTDPAVKNWMTQTLLIQNANPDCKLV
                     LKGLGMNPTLEEMLTACQGVGGPGQKARLMAEALKEALAPGPLPFAAAQQKGPRKPIK
                     CWNCGKEGHSARQCRTPRRQGCWKCGQMGHVMAKCPDRQAGFLGFGPWGKKPRNFPMA
                     QMHQGLTPTAPPEDPAVDLLKNYMQLGKQQRESKRKPYKEVTEDLLHLNSLFGEDQ"
CDS                  1708. .4878
                     /partial
                     /note="NH2-terminus uncertain"
                     /codon_start=1
                     /product="pol polyprotein"
                     /protein_id="AAA91931.1"
                     /db_xref="GI:334697"
                     /translation="VLELWERGTLCKAMQNPKKTGLLEMWTNGPCYGQMPRQTGRFFR
                     LWPMGKEAPQFPHGPNASGADANCSPRGPSCGSAKKLHAVGQTAERKQKEALQGGDRG
                     FAAPQFSLWRRPVVTAHIEGQPAEVLLDTGADDSIVAGIELGPHYTPKIVGGIGGFIN
                     TKEYKNVKIEVLGKRIKGTIMTGDTPINIFGRNLLTALEMSLNFPIAKVEPVKVTLKP
                     GKDGPKLRQWPLSKEKIVALREICEKMEKDGQLEEAPPTNPYNTPTFAIKKKDKNKWR
                     ILIDFRELNKVTQDFTEVQLGIPHPAGLAKRKRITVLDVGDAYFSIPLDEEFRQYTAF
                     TLPSVNNAEPGKRYIYKVLPQGWKGSPAIFQHTMRNVLEPFRKANPDVTLVQYMDDIL
                     VASDRTDLEHDRVVLQLKELLNSIGFSTPEEKFQKDPPFQWMGYELWPTKWKLQKIEL
                     PQKETWTVNDIQKLVGVLNWAAQIYPGIKTKHLCRLIRGKMTLTEEVQWTEMAEAEYE
                     ENKIILSQEQEGCYYQEGKPLEATVIKNQDNQWSYKIHQEDKILKVGKFAKIKNTHTN
                     GVRLLAHVIQKIGKEAIVIWGQVPKFHLPVEKDVWEQWWTDYWQVTWIPKWDFISTPP
                     LVRLVFNLVKDPIKGEETYYVDGSCNKQSKEGKAGYITDRGKDKVKVLEQTTNQQAEL
                     EAFLMALADSGPKANIIVDSQYVMGIITGCPTESESRLVNQIIEEMIKKTEIYVAWVP
                     AHKGIGGNQEIDHLVSQGIRQVLFLEKIEPAQEEHDKYHSNVKELVFKFGLPRLVAKQ
                     IVDTCDKCHQKGEAIHGQVNSDLGTWQMDCTHLEGKIIIVAVHVASGFIEAEVIPQET
                     GRQTALFLLKLASRWPITHLHTDNGANFASQEVKMVAWWAGIEHTFGVPYNPQSQGVV
                     EAMNHHLKNQIDRIREQANSMETIVLMAVHCMNFKRRGGIGDMTPAERLLNMITTEQE
                     IQFQQSKNSKFKNFRVYYREGRDQLWKGPGELLWKGEGAVVLKVGTDIKVVPRRKAKI
                     IKDYGGGKEVDSSSHMEDTGEAREVA"
CDS                  4808. .5452
                     /note="vif protein"
                     /codon_start=1
                     /protein_id="AAA91932.1"
                     /db_xref="GI:334698"
                     /translation="MEEEKRWIAVPTWRIPERLERWHSLIKYLKYKTKDLQKVCYVPH
                     HKVGWAWWTCSRVIFPLQEKSQLEVQGYWNLTPERGWLSTHAVRITWYSRNFWTDVTP
                     DCADILLHSTYFPCFTAGEVRRAIRGEQLLSCCRFPRAHKTQVPSLQYLALRVVSYVR
                     SQRENPTWKQWRRDNRRSLRMAKQNSRGDKQRGSKPPTKGVDFPGLAKVLGILA"
CDS                  5280. .5618
                     /note="vpX protein"
                     /codon_start=1
                     /protein_id="AAA91933.1"
                     /db_xref="GI:334699"
                     /translation="MSDPRERIPPGNSGEETIGEAFEWLNRTVEEINREAVNHLPREL
                     IFQVWQRSWEYWHDEQGMSQSYVKYRYLCLIQKALFMHCKKGCRCLGEGHGAGGWRPG
                     PPPPPPPGLA"
CDS                  5619. .5924
                     /note="vpR protein"
                     /codon_start=1
                     /protein_id="AAA91934.1"
                     /db_xref="GI:334700"
                     /translation="MEERPPEDEGPQREPWDEWVVEVLEELKEEALKHFDPRLLTALG
                     NHIYNRHGDTLEGAGELIKILQRALFMHFRGGCTHSRIGQSGGGNPLSTIPPSRRML"
CDS                  join(5770. .6065,8280. .8376)
                     /note="tat protein"
                     /codon_start=1
                     /protein_id="AAA91935.1"
                     /db_xref="GI:334693"
                     /translation="METPLREQENSLKSSNGRSSCTSEAAAPTLESANLEEEILSQLY
                     RPLEECYNTCYCKKCCYHCQFCFLKKGLGICYEQSRRRRRTPKKAKANTSSASNKPIS
                     DRTRHCQPKKEQKETVEAAVATAPGLGR"
CDS                  join(5996. .6065,8280. .8533)
                     /note="rev protein"
                     /codon_start=1
                     /protein_id="AAA91936.1"
                     /db_xref="GI:334694"
                     /translation="MSSHAEEEELRRRLRLIHLLHQTNPYPTGPGTANQRRNRRRRWR
                     QRWQQLLALADRIYSFPDPPTNTPLDLVIQQLQNLAIESIPDPPTNIPEILHDPTENP
                     RSPQD"
CDS                  6072. .8777
                     /note="env protein (premature stop codon)"
                     /pseudo
                     /codon_start=1
CDS                  join(6072. .8276,8280. .8717)
                     /note="env protein"
                     /codon_start=1
                     /protein_id="AAA91937.1"
                     /db_xref="GI:334695"
                     /translation="MGCLGNQLLIAILFLSAYGIYCIQYVTVFYGVPAWRNATIPLFC
                     VTRNRDTWGTTQCLPDNDDYSELALNITESFDAWENTVTEQAIEDVWHLFETSIKPCV
                     KLTPLCITMKCNKSETDKWGLTKSSTTTAPTAIPTKAEAIKVVNENSPCINHDNCTGL
                     EQEPMISCKFNMTGLKRDKRREYNETWYSADLVCEQGNSTENESRCYMNHCNTSVIQE
                     SCDKHYWDAIRFRYCAPPGYALLRCNDTNYSGFMPNCSKVVVSSCTRMMETQTSTWFG
                     FNGTRAENRTYIYWHSKDNRTIISLNKYYNLTMKCRRPGNKTVLPVTIMSGLVFHSQP
                     INDRPKQAWCRFEGNWKEAIKEVKQTIVKHPRYTGTNNTDKINLTAPGGGDPEVTFMW
                     TNCRGEFLYCKMNWFLNWVEDKNLTGTTQKPQERHKRNYVPCHIRQIINTWHKVGRNV
                     YLPPREGDLTCNSTVTSLIANIDWIDGNQTNITMSAEVAELYRLELGDYKLVEITPIG
                     LAPTNVKRYTTGGTSRNKRGVFVLGFLGFLATAGSAMGAASLTLTAQSRTLLAGIVQQ
                     QQQLLDVVKRQQELLRLTVWGTKNLQTRVTAIEKYLKDQAQLNAWGCAFRQVCHTTVP
                     WPNANLTPNWNNETWQEWERKVDFLEENITALLEEAQIQQEKNMYELQKLNSWDVFGN
                     WFDLASWIRYIQYGVYIVVGVILLRIVIYIVQMLAKLRQGYRPVFSSPPSYFQTHIRQ
                     DQALPTKEGTEGDGGGSGGNSSWPWQIEYIHFLIRQLIRLLTWLFSNCRTLLSRAYQI
                     LQPIFQRFSTTLQRIREVLRTELTYLQYGWSYFQEAVQVAWRSATETLAGAWGDLWET
                     LGRVGRWILAIPRRIRQGLELTLL"
misc_feature         8277. .8279
                     /note="env protein in-frame stop codon"
CDS                  8551. .9342
                     /note="nef protein"
                     /codon_start=1
                     /protein_id="AAA91938.1"
                     /db_xref="GI:334701"
                     /translation="MGGATSKRRSKSPGDLRQRLLRARGETYGRLWEELEDGYSQSLG
                     GSDKGSSLLSCEGQKYNQGQFMNTPWKNPAGEREKLAYRKQNIDDIDEEDNDLVGVPV
                     RPRVPLRIISYKLAVDMSHFIKEKGGLEGIYYSERRHKILDMYLEKEEGIMPDWQNYT
                     SGPGPRYPKTFGWLWKLVPVNVSDEAQEGEENYLLHPAQTSQWDDPWGEVLVWKFDPT
                     LAYTYEAYIRYPEEFGSKSGLSEEEVRRRLTARGLLKMADKRETS"
BASE COUNT     3294 a   1807 c   2379 g   2148 t
ORIGIN      5' end of 5' LTR R region (putative mRNA start).

  Query Match             95.7%;  Score 2453.6;  DB 14;  Length 9628;
  Best Local Similarity   97.3%;  Pred. No. 0;
  Matches 2495;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy       1 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 60
           || |||||||||| |||||||||||||||||| |||||||||||||||| |||||||||
Db    2317 TTCCCCATAGCTAAGGTAGAGCCTGTAAAAGTCACCTTAAAGCCAGGAAAAGATGGACCA 2376

Qy      61 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 120
           ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2377 AAATTGAGGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 2436

Qy     121 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2437 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 2496

Qy     181 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 240
           |||||||| |||||||||||||| |||||||||||||||| ||||||||||||||||||
Db    2497 ACATTTGCCATAAAGAAAAAGGACAAGAACAAATGGAGAATACTGATAGATTTTAGGGAA 2556

Qy     241 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 300
           ||||||| |||||||||||||||| |||||||||| |||||||||||||||||||||||
Db    2557 CTAAATAAGGTCACTCAGGACTTTACAGAAGTCCAATTGGGAATACCACACCCTGCAGGA 2616

Qy     301 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 360
           |||||||||||||| || || |||||||||||| |||||||| ||||||||||||||||||
Db    2617 CTAGCAAAAAGGAAGAGGATCACAGTACTGGATGTAGGTGACGCATATTTCTCCATACCT 2676

Qy     361 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 420
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
```


Db 2677 CTAGATGAAGAATTTAGGCAGTACACTGCTTTTACTTTACCATCAGTAAATAATGCAGAA 2736
Qy 421 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 480
Db 2737 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGGTGGAAGGGGTCACCAGCCATC 2796
Qy 481 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 540
Db 2797 TTCCAACACACTATGAGAAATGTGCTGGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 2856
Qy 541 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 600
Db 2857 TTAGTCCAGTATATGGATGACATCTTAGTAGCTAGTGACAGGACAGACCTGGAACATGAC 2916
Qy 601 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 660
Db 2917 AGGGTAGTTTTACAGTTAAAGGAACTCTTAAATAGCATAGGGTTTTCTACCCCAGAAGAG 2976
Qy 661 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 720
Db 2977 AAGTTCCAAAAAGATCCCCCATTTCAATGGATGGGGTATGAATTGTGGCCAACAAAATGG 3036
Qy 721 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 780
Db 3037 AAGTTGCAAAAGATAGAGTTGCCACAAAAAGAGACCTGGACAGTGAATGATATACAGAAG 3096
Qy 781 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 840
Db 3097 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 3156
Qy 841 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 900
Db 3157 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 3216
Qy 901 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 960
Db 3217 GAGGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 3276
Qy 961 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 1020
Db 3277 CAAGAAGGCAAGCCATTAGAGGCCACGGTAATAAAGAATCAGGACAATCAGTGGTCTTAT 3336
Qy 1021 AAAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACA 1080
Db 3337 AAGATTCACCAAGAAGACAAAATACTAAAAGTAGGAAAATTTGCAAAGATAAAAAATACA 3396
Qy 1081 CATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATA 1140
Db 3397 CATACCAATGGAGTTAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATA 3456
Qy 1141 GTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAG 1200
Db 3457 GTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAAGATGTATGGGAACAG 3516
Qy 1201 TGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCA 1260
Db 3517 TGGTGGACAGACTATTGGCAGGTAACCTGGATACCGAAATGGGATTTTATCTCAACACCA 3576
Qy 1261 CCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTAT 1320
Db 3577 CCACTAGTAAGATTAGTCTTCAATCTGGTAAAGGACCCTATAAAGGGAGAAGAAACCTAT 3636
Qy 1321 TATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGAT 1380
Db 3637 TATGTAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGAT 3696
Qy 1381 AGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAA 1440
Db 3697 AGGGGCAAAGACAAAGTAAAAGTCTTAGAACAGACTACTAATCAACAAGCAGAATTGGAA 3756
Qy 1441 GCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAA 1500
Db 3757 GCATTTCTCATGGCATTGGCAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAA 3816
Qy 1501 TATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAA 1560
Db 3817 TATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAACCAA 3876
Qy 1561 ATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAA 1620
Db 3877 ATAATAGAAGAAATGATTAAAAAGACAGAAATTTATGTAGCATGGGTGCCAGCACACAAA 3936
Qy 1621 GGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTC 1680
Db 3937 GGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTC 3996
Qy 1681 TTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAA 1740
Db 3997 TTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAA 4056
Qy 1741 GAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGT 1800
Db 4057 GAATTGGTATTCAAATTTGGATTACCCAGACTAGTGGCCAAACAGATAGTAGACACATGT 4116
Qy 1801 GATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACT 1860
Db 4117 GATAAATGTCATCAGAAAGGAGAAGCTATACATGGGCAGGTAAATTCAGATCTAGGGACT 4176
Qy 1861 TGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCT 1920
Db 4177 TGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCT 4236
Qy 1921 AGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTT 1980
Db 4237 AGTGGATTCATAGAAGCAGAAGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTT 4296
Qy 1981 CTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAAC 2040
Db 4297 CTGTTAAAATTGGCAAGCAGATGGCCTATTACGCATCTACACACAGATAATGGTGCCAAC 4356
Qy 2041 TTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGG 2100
Db 4357 TTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGG 4416
Qy 2101 GTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAAT 2160
Db 4417 GTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCATCTAAAAAAT 4476
Qy 2161 CAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTT 2220
Db 4477 CAAATAGATAGAATCAGGGAACAAGCAAATTCAATGGAAACCATAGTATTAATGGCAGTT 4536
Qy 2221 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTA 2280
Db 4537 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTA 4596
Qy 2281 ATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTT 2340
Db 4597 CTTAACATGATCACTACAGAACAAGAAATACAATTCCAACAATCAAAAAACTCAAAATTT 4656
Qy 2341 AAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAG 2400
Db 4657 AAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAGCTGTGGAAAGGACCTGGTGAG 4716
Qy 2401 CTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTA 2460
Db 4717 CTATTGTGGAAAGGGGAAGGAGCAGTCGTCTTAAAGGTAGGGACAGACATTAAGGTAGTA 4776
Qy 2461 CCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGT 2520
Db 4777 CCCAGAAGAAAGGCTAAGATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGT 4836
Qy 2521 TCCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 2564
Db 4837 TCCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 4880

```
RESULT 15
SIU72748
LOCUS       SIU72748               10289 bp    DNA     linear   VRL 06-FEB-1997
DEFINITION  Simian immunodeficiency virus strain SIVsmE543, complete genome.
ACCESSION   U72748
VERSION     U72748.1  GI:1695908
KEYWORDS    .
SOURCE      Simian immunodeficiency virus.
  ORGANISM  Simian immunodeficiency virus
            Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
            lentivirus group.
REFERENCE   1  (bases 1 to 10289)
  AUTHORS   Hirsch,V., Adger-Johnson,D., Campbell,B., Goldstein,S., Brown,C.,
            Elkins,W.R. and Montefiori,D.C.
  TITLE     A molecularly cloned, pathogenic, neutralization-resistant simian
            immunodeficiency virus, SIVsmE543-3
  JOURNAL   J. Virol. 71 (2), 1608-1620 (1997)
  MEDLINE   97151152
   PUBMED   8995688
REFERENCE   2  (bases 1 to 10289)
  AUTHORS   Hirsch,V.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-1996) Laboratory of Infectious Diseases, National
            Institutes of Health, NIAID, Twinbrook II, 12441 Parklawn Drive,
            Rockville, MD 20852, USA
FEATURES             Location/Qualifiers
     source          1. .10289
                     /organism="Simian immunodeficiency virus"
                     /proviral
                     /strain="SIVsmE543"
                     /db_xref="taxon:11723"
                     /clone="SIVsmE543-3"
                     /cell_line="CEMx174"
                     /cell_type="lymphocyte"
                     /lab_host="Macaca mulatta"
     LTR             1. .817
     gene            1050. .2573
                     /gene="gag"
     CDS             1050. .2573
                     /gene="gag"
                     /codon_start=1
                     /product="gag protein"
                     /protein_id="AAC56558.1"
                     /db_xref="GI:1695909"
                     /translation="MGARNSVLSGKEADELEKVRLRPNGKKKYMLKHVVWAANELDRF
                     GLAESLLDNKEGCQKILSVLAPLVPTGSENLKSLYNTVCVIWCIHAEEKVKHTEEAKQ
                     IVQRHLVVETGTADKMPATSRPTAPPSGRGGNYPVQQVGGNYVHLPLSPRTLNAWVKL
                     VEEKKFGAEVVPGFQALSEGCTPYDINQMLNCVGEHQAAMQIIREIINEEAADWDLQH
                     PQPGPLPAGQLREPRGSDIAGTTSTVEEQIQWMYRQQNPIPVGNIYRRWIQLGLQKCV
                     RMYNPTNILDVKQGPKEPFQSYVDRFYKSLRAEQTDPAVKNWMTQTLLIQNANPDCKL
                     VLKGLGMNPTLEEMLTACQGIGGPGQKARLMAEALKEALRPDQLPFAAVQQKGQRRTI
                     KCWNCGKEGHSARQCRAPRRQGCWGCGKTGHVMAKCPERQAGFLGFGPWGKKPRNFPM
                     AQMPQGLIPTAPPEDPAVDLLKNYMKMGRKQRENRERPYKEVTEDLLHLNSLFGEDQ"
     gene            <2339. .5398
                     /gene="pol"
     CDS             <2339. .5398
                     /gene="pol"
                     /codon_start=1
                     /product="pol polyprotein"
                     /protein_id="AAC56559.1"
                     /db_xref="GI:1695910"
                     /translation="KTGGFFRVWPMGKEAPQFPHGPDASGADTNCSPRGSSCGSTEEL
                     HEDGQKAEGEQRETLQGGDRGFAAPQFSLWRRPVVTAYIEEQPVEVLLDTGADDSIVT
                     GIELGPNYTPKIVGGIGGFINTKEYKDVKIKVLGKVIKGTIMTGDTPINIFGRNLLTA
                     MGMSLNFPIAKVEPIKVTLKPGKEGPKLRQWPLSKEKIIALREICEKMEKDGQLEEAP
                     PTNPYNTPTFAIKKKDKNKWRMLIDFRELNKVTQDFTEVQLGIPHPAGLAKRRRITVL
                     DVGDAYFSIPLDEEFRQYTAFTLPSVNNAEPGKRYIYKVLPQGWKGSPAIFQYTMRNV
                     LEPFRKANPDVTLIHYMDDILIASDRTDLEHDRVVLQLKELLNGIGFSTPEEKFQKDP
                     PFQWMGYELWPTKWKLQKIELPQRETWTVNDIQKLVGVLNWAAQIYPGIKTKHLCRLI
                     RGKMTLTEEVQWTEMAEAEYEENKIILSQEQEGCYYQEGKPIEATVIKSQDNQWSYKI
                     HQEDKVLKVGKFAKVKNTHTNGVRLLAHVVQKIGKEALVIWGEVPKFHLPVEREIWEQ
                     WWTDYWQVTWIPDWDFVSTPPLVRLVFNLVKEPIQGAETFYVDGSCNRQSREGKAGYV
                     TDRGRDKAKLLEQTTNQQAELEAFYLALADSGPKANIIVDSQYVMGIVAGQPTESESR
                     LVNQIIEEMIKKEAIYVAWVPAHKGIGGNQEVDHLVSQGIRQVLFLEKIEPAQEEHEK
                     YHSNVKELVFKFGIPRLVAKQIVDTCDRCHQKGEAIHGQVNAELGTWQMDCTHLEGKI
                     IIVAVHVASGFIEAEVIPQETGRQTALFLLKLAGRWPITHLHTDNGANFTSQEVKMVA
                     WWAGIEQTFGVPYNPQSQGVVEAMNHHLKTQIDRIREQANSVETIVLMAVHCMNFKRR
                     GGIGDMTPAERLVNMITTEQEIQFQQSKNSKFKNFRVYYREGRDQLWRGPGELLWKGE
                     GAVILKVGTEIKVVPRRKAKIIKDYGGGKELDSGSHLEDTGEAREVA"
     gene            5328. .5972
                     /gene="vif"
     CDS             5328. .5972
                     /gene="vif"
                     /codon_start=1
                     /protein_id="AAC56560.1"
                     /db_xref="GI:1695911"
                     /translation="MEEEKNWIVVPTWRIPERLERWHSLIKHLKYNTKDLQMACYVPH
                     HKVGWAWWTCSRVIFPLRDETHLEVQGYWNLAPEKGWLSTHAVRITWYSRNFWTDVTP
                     DCADTLLHSTYFPCFSEGEVQRAIRGEKLLSCCKFPKAHKNQVPSLQYLALTVVSHVR
                     SQREDPTWKQWRGNNERGLRMAKQNSRRNKQGSSKSPAEGANFPGLAKVLGILA"
     gene            5800. .6138
                     /gene="vpx"
     CDS             5800. .6138
                     /gene="vpx"
                     /codon_start=1
                     /protein_id="AAC56561.1"
                     /db_xref="GI:1695912"
                     /translation="MSDPRERIPPGNSGEETIEEAFEWLNRTVEGINRAAVNHLPREL
                     IFQVWQRSWEYWHDEMGMSESYTKYRYLILIQKALFMHCKKGCRCLGEGHGAGGWRTG
                     PPPPPPPGLA"
     gene            6140. .6445
                     /gene="vpr"
     CDS             6140. .6445
                     /gene="vpr"
                     /codon_start=1
                     /protein_id="AAC56562.1"
                     /db_xref="GI:1695913"
                     /translation="MAERPPEDEAPQREPWDEWVVEVLEEIKEEALKHFDPRLLTALG
                     NYIYDRHGDTLEGAGELIRILQRALFIHFRSGCAHSRIGQSRGGNPLSTIPPSRAML"
     gene            6282. .8909
                     /gene="tat"
     CDS             join(6282. .6580,8810. .8909)
                     /gene="tat"
                     /codon_start=1
                     /product="tat protein"
                     /protein_id="AAC56563.1"
                     /db_xref="GI:1695914"
                     /translation="MIDMETPLKEQENSLESYREHSSSISEVDVPTPESANLEEEILS
                     QLYRPLEPCYNKCYCKRCCYHCQHCFLKKGLGICYEQHRRRTPKKTKTNPLPASNNRS
                     LSTRTRNRQPKKEKKEKVETEVAADLGLGR"
     gene            6517. .9051
                     /gene="rev"
     CDS             join(6517. .6580,8810. .9051)
                     /gene="rev"
                     /codon_start=1
                     /product="rev protein"
                     /protein_id="AAC56564.1"
                     /db_xref="GI:1695915"
                     /translation="MSSTEEELRKRLRLIHFLHQTTDPYPQGPGTANQRRRRRRRWRQ
                     RWQQILALADRIYSFPNPPTDTPLDLAIQQLQGLAIEDLPDPPTSAPETLKDAAKSS"
     gene            6587. .9250
                     /gene="env"
     CDS             6587. .9250
                     /gene="env"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAC56565.1"
                     /db_xref="GI:1695916"
                     /translation="MGCLGNQLLIALLLVSVLEICCVQYVTVFYGVPAWKNATIPLFC
                     ATRNRDTWGTTQCLPDNDDYSELAVNITEAFDAWNNTVTEQAIEDVWNLFETSIKPCV
                     KLTPLCIAMRCNKTETDRWGLTGRAETTTTAKSTTSTTTTTVTPKVINEGDSCIKNNS
                     CAGLEQEPMIGCKFNMTGLKRDKKIEYNETWYSRDLICEQPANGSESKCYMQHCNTSV
                     IQESCDKHYWDAIRFRYCAPPGYALLRCNDSNYLGFAPKCSKVVVSSCTRMMETQTST
                     WFGFNGTRAENRTYIYWHGNSNRTIISLNKYYNLTMKCRRPGNKTVLPVTIMSGLVFH
                     SQPINERPKQAWCRFGGNWSEAIQEVKETLVKHPRYTGTNDTRKINLTAPAGGDPEVT
```

```
                     FMWTNCRGEFLSCKMNWFLNWVEDRDQNSNRWKQQKKPEQQKRNYVPCHIRQIINTWH
                     KVGKNVYLPPREGDLTCNSTVTSLIAEIDWINNNETNITMSAEVAELYRLELGDYKLV
                     EITPIGLAPTDVRRYTTTGASRNKRGVFVLGFLGFLATGGSAMGAASLTVSAQSRTLL
                     AGIVQQQQQLLDVVKRQHELLRLTVWGTKNLQTRVTAIEKYLKDQAQLNSWGCAFRQV
                     CHTTVPWPNDSLVPNWDNMTWQEWEGKVDFLEANITQLLEEAQIQQEKNMYELQKLNS
                     WDIFGNWFDLTSWIRYIQYGVLIVLGVVGLRIVIYVVQMLARLRQGYRPVFSPPPAYV
                     QQIPIHKDQEPPTKEGEEGEGGDRGGSRSWPWQIEYIHFLIRQLIRLLTWLFSSCRDW
                     LLRIYQILQPVLQRLSRTLQRVREVIRIEIITYLQYGWSYFQEAVQAWWKFARETLASA
                     WRDIWETLGRVGRGILAIPRRVRQGLELALL"
     gene            9084. .9863
                     /gene="nef"
     CDS             9084. .9863
                     /gene="nef"
                     /codon_start=1
                     /product="nef protein"
                     /protein_id="AAC56566.1"
                     /db_xref="GI:1695917"
                     /translation="MGGAISKKQYRRGGNLRERLLRARGETYGRLWEGLEEGYSQSLG
                     ASGKGLSSLSCEPQKYSEGQYMNTPWRNPTAEKAKLGYKQQNMDDVDDEDDDLVGVSV
                     HPKVPLRAMTYKLAIDMSHFIKEKGGLEGIYYNEKRHRILDMYMEKEEGIIPDWQNYT
                     LGPGTRYPMHFGWLWKLVPVDVSDEAQEDETHCLVHPAQTHQWDDPWGEVLAWKFDPE
                     LAYSYKAFIKYPEEFGSKSGLSEEEVKRRLTARGIYKMADK"
BASE COUNT     3488 a   1915 c   2583 g   2303 t
ORIGIN

  Query Match             78.2%;  Score 2004;  DB 14;  Length 10289;
  Best Local Similarity   86.3%;  Pred. No. 0;
  Matches 2214;  Conservative    0;  Mismatches  350;  Indels    0;  Gaps    0;

Qy       1 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 60
Db    2837 TTCCCCATAGCTAAGGTGGAGCCTATAAAAGTAACACTAAAACCAGGGAAAGAAGGACCA 2896

Qy      61 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 120
Db    2897 AAATTGAGACAGTGGCCGCTATCAAAAGAAAAGATAATTGCATTAAGAGAAATCTGTGAA 2956

Qy     121 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 180
Db    2957 AAAATGGAAAAAGATGGCCAGTTAGAGGAAGCCCCTCCAACCAATCCGTATAACACCCCC 3016

Qy     181 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 240
Db    3017 ACTTTTGCTATAAAGAAGAAAGACAAAAATAAATGGAGAATGCTAATAGATTTTAGAGAA 3076

Qy     241 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 300
Db    3077 TTAAATAAGGTCACTCAAGACTTTACAGAAGTACAGTTAGGAATACCACACCCTGCAGGA 3136

Qy     301 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 360
Db    3137 CTAGCAAAGAGAAGGAGGATCACAGTATTGGATGTAGGTGATGCATATTTCTCCATTCCT 3196

Qy     361 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 420
Db    3197 CTAGATGAAGAATTCAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAA 3256

Qy     421 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 480
Db    3257 CCAGGAAAAAGATACATCTATAAGGTATTACCTCAAGGGTGGAAGGGGTCACCAGCTATT 3316

Qy     481 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 540
Db    3317 TTTCAGTATACTATGAGAAATGTATTAGAACCTTTCAGAAAAGCAAATCCAGATGTGACC 3376

Qy     541 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 600
Db    3377 CTGATCCACTACATGGATGACATCTTAATAGCTAGTGATAGAACAGATTTAGAGCATGAC 3436

Qy     601 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 660
Db    3437 AGGGTAGTTTTACAGTTAAAGGAACTTCTGAACGGCATAGGATTCTCCACCCCAGAAGAG 3496

Qy     661 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 720
Db    3497 AAGTTCCAGAAAGATCCCCCATTCCAGTGGATGGGATATGAATTGTGGCCAACCAAATGG 3556

Qy     721 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 780
Db    3557 AAACTGCAGAAAATAGAGTTGCCACAAAGAGAGACCTGGACAGTAAATGACATACAAAAA 3616

Qy     781 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 840
Db    3617 TTAGTAGGAGTGCTAAATTGGGCAGCACAAATTTATCCAGGAATAAAGACTAAACATCTT 3676

Qy     841 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 900
Db    3677 TGCAGACTAATCAGAGGAAAAATGACTTTAACAGAAGAGGTTCAGTGGACTGAGATGGCA 3736

Qy     901 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 960
Db    3737 GAGGCAGAATATGAAGAAAACAAGATAATTCTCAGTCAAGAACAAGAAGGATGTTACTAC 3796

Qy     961 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 1020
Db    3797 CAAGAGGGAAAACCAATAGAGGCAACAGTAATAAAGAGTCAGGATAATCAGTGGTCATAT 3856

Qy    1021 AAAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACA 1080
Db    3857 AAAATTCACCAAGAAGACAAAGTACTAAAAGTGGGCAAATTTGCAAAGGTTAAAAATACA 3916

Qy    1081 CATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATA 1140
Db    3917 CATACAAATGGAGTCAGATTACTAGCACACGTAGTGCAGAAAATAGGAAAGGAGGCACTA 3976

Qy    1141 GTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAG 1200
Db    3977 GTAATTTGGGGAGAGGTGCCAAAATTCCATTTGCCAGTAGAAAGAGAAATTTGGGAACAA 4036

Qy    1201 TGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCA 1260
Db    4037 TGGTGGACAGATTATTGGCAAGTAACCTGGATACCAGATTGGGACTTTGTGTCAACACCT 4096

Qy    1261 CCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTAT 1320
Db    4097 CCCTTAGTCAGATTAGTCTTCAACCTAGTAAAAGAGCCTATACAGGGAGCAGAAACATTT 4156

Qy    1321 TATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGAT 1380
Db    4157 TATGTAGATGGATCCTGTAATAGGCAGTCAAGAGAAGGAAAAGCAGGCTATGTGACGGAT 4216

Qy    1381 AGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAA 1440
Db    4217 AGGGGCAGAGACAAAGCAAAACTTTTAGAACAGACTACCAACCAACAAGCAGAGTTGGAA 4276

Qy    1441 GCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAA 1500
Db    4277 GCCTTCTATCTAGCCTTAGCAGATTCGGGACCAAAAGCAAATATTATAGTAGATTCCCAA 4336

Qy    1501 TATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAA 1560
Db    4337 TATGTTATGGGCATAGTAGCAGGTCAGCCCACTGAATCAGAAAGTAGGTTAGTAAACCAG 4396

Qy    1561 ATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAA 1620
Db    4397 ATAATAGAAGAGATGATTAAAAAGGAAGCAATTTATGTAGCATGGGTACCTGCACATAAA 4456

Qy    1621 GGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTC 1680
Db    4457 GGAATAGGAGGAAATCAAGAAGTAGATCACCTGGTTAGCCAGGGAATTAGACAAGTCCTA 4516

Qy    1681 TTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAA 1740
Db    4517 TTCTTAGAAAAAATAGAACCAGCACAAGAAGAGCATGAAAAGTACCATAGCAATGTAAAA 4576

Qy    1741 GAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGT 1800
```

```
Db   4577 GAATTGGTATTCAAATTTGGTATACCTAGGCTAGTAGCAAAACAGATAGTAGACACATGT 4636

Qy   1801 GATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACT 1860

Db   4637 GATAGATGCCACCAGAAAGGAGAAGCCATACATGGGCAAGTAAATGCAGAACTAGGGACT 4696

Qy   1861 TGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCT 1920

Db   4697 TGGCAAATGGACTGTACGCACCTAGAAGGCAAAATAATTATAGTTGCAGTACATGTGGCT 4756

Qy   1921 AGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTT 1980

Db   4757 AGTGGATTTATAGAGGCAGAAGTAATCCCGCAGGAAACAGGAAGACAAACAGCACTGTTT 4816

Qy   1981 CTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAAC 2040

Db   4817 CTGTTAAAGCTAGCTGGCAGATGGCCCATCACACATCTGCATACTGATAATGGTGCCAAT 4876

Qy   2041 TTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGG 2100

Db   4877 TTCACATCACAAGAAGTGAAAATGGTTGCCTGGTGGGCAGGGATTGAACAGACCTTTGGA 4936

Qy   2101 GTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAAT 2160

Db   4937 GTGCCTTATAATCCACAGAGCCAAGGAGTAGTGGAAGCAATGAACCATCATTTAAAAACC 4996

Qy   2161 CAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTT 2220

Db   4997 CAGATAGATAGAATTAGAGAACAAGCAAACTCAGTAGAGACTATAGTACTAATGGCAGTT 5056

Qy   2221 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTA 2280

Db   5057 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTA 5116

Qy   2281 ATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTT 2340

Db   5117 GTCAATATGATCACCACAGAACAAGAAATACAATTCCAACAATCAAAAAATTCAAAATTT 5176

Qy   2341 AAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAG 2400

Db   5177 AAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGACCAGCTGTGGAGAGGACCCGGTGAG 5236

Qy   2401 CTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTA 2460

Db   5237 CTATTGTGGAAAGGGGAAGGAGCAGTCATCCTAAAGGTAGGGACAGAGATTAAGGTAGTA 5296

Qy   2461 CCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGT 2520

Db   5297 CCAAGGAGGAAAGCTAAAATTATCAAAGACTATGGAGGAGGAAAAGAATTGGATAGTGGT 5356

Qy   2521 TCCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 2564

Db   5357 TCCCACTTGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 5400
```

Search completed: November 16, 2002, 11:21:59
Job time : 4658.28 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 16, 2002, 06:33:28 ; Search time 353.009 Seconds
(without alignments)
16356.861 Million cell updates/sec

Title: M33262_COPY_3105_5668
Perfect score: 2564
Sequence: 1 TTTCCCATAGCTAAAGTAGA.....GCTAGAGAGGTGGCATAGCC 2564

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2564	100.0	10006	24	AAD25514	Simian-Human immun
2	2564	100.0	10279	20	AAV81865	Nucleic acid seque
3	2562.4	99.9	10097	13	AAQ24802	SIVmac239 nef-dele
4	2562.4	99.9	10279	13	AAQ22487	SIVmac239 proviral
5	2510.2	97.9	10274	24	AAD25516	Simian-Human immun
6	2467	96.2	9601	9	AAN80860	Sequence corresp.
7	1751.8	68.3	10163	18	AAT61085	Full-length HIV-2K
8	1738.4	67.8	9693	10	AAN92119	Sequence of clone
9	1734.2	67.6	9663	21	AAC81938	Packaging vector p
10	1734.2	67.6	9726	21	AAC81937	Packaging vector p
11	1727.8	67.4	9672	13	AAQ20616	ROD HIV-2 isolate
12	1724.2	67.2	9633	9	AAN80890	Sequence of cDNA c
13	1723	67.2	9822	10	AAN91774	Entire HIV-2/ST pr
14	1721	67.1	9630	21	AAZ89628	HIV-2 ROD DNA. Hu
15	1717.6	67.0	9360	11	AAQ02830	cDNA to HIV-2 RNA.
16	1711.8	66.8	9473	10	AAN92768	HIV-2 variant HIV-
17	1695.2	66.1	3168	21	AAA87840	HIV-2 reverse tran
18	1671.8	65.2	9643	9	AAN80859	Sequence of entire
19	1587	61.9	2689	20	AAX05230	HIV-2 genomic DNA
20	1497.6	58.4	4358	24	AAI67895	Nucleotide sequenc
21	1416	55.2	7984	24	ABK91610	Modified HIV prote
22	1416	55.2	9280	24	ABK91615	Modified HIV prote
23	1416	55.2	9285	24	ABK91608	Modified HIV prote
24	1343.4	52.4	4414	10	AAN92771	HIV-2 variant HIV-
25	1341.8	52.3	4414	16	AAQ92470	HIV-D205 partial s
26	1196	46.6	9641	22	AAH77486	Mandrill immunodef
27	1171.8	45.7	9465	21	AAA51008	Simian immunodefic
28	1089.8	42.5	9745	20	AAX04766	The complete HIV-1
29	1089.8	42.5	9746	18	AAT58550	Human Immunodefici
30	1089.4	42.5	9215	12	AAQ10203	Sequence of simian
31	1086.2	42.4	3015	24	ABL39984	Wild type 8_5_TV1_
32	1086.2	42.4	9781	24	ABL39962	8_5_TV1_C.ZA nucleo
33	1083.4	42.3	9425	14	AAQ45921	HTLV-III DNA (Clon
34	1080.2	42.1	9143	11	AAQ06635	Complete sequence
35	1076.6	42.0	3009	24	ABL40025	Wild type 12-5_1_T
36	1076.6	42.0	9738	24	ABL39966	12-5_1_TV2_C.ZA nu
37	1075.4	41.9	2709	21	AAA87842	HIV-1 reverse tran
38	1075.4	41.9	5340	8	AAN71017	Sequence of LAV/HT
39	1075	41.9	8960	21	AAF57929	HIV-1 non-subtype
40	1073.8	41.9	9193	7	AAN60365	Sequence of LAV vi
41	1073.8	41.9	9609	24	AAD25515	Human immunodefici
42	1072.2	41.8	8992	21	AAF57922	HIV-1 non-subtype
43	1072.2	41.8	9196	20	AAX03987	HIV-1-NC infectiou
44	1072.2	41.8	9229	19	AAV63467	HIV isolate LAV.MA
45	1071.8	41.8	9078	22	AAH20867	HIV-1 subtype (C/B

ALIGNMENTS

RESULT 1
AAD25514
ID AAD25514 standard; DNA; 10006 BP.
XX
AC AAD25514;
XX
DT 26-MAR-2002 (first entry)
XX
DE Simian-Human immunodeficiency virus clone SHIV-HXBcP 3.2 complete genome.
XX
KW HIV; human immunodeficiency virus; simian virus; cytostatic; cancer;
KW immunosuppressive; virucide; antibacterial; fungicide; protozoacide;
KW antirheumatic; antiinflammatory; antiarthritic; rheumatoid arthritis;
KW neuroprotective; multiple sclerosis; immune response; vasotropic;
KW vaccine; gene therapy; autoimmune disease; vasculitis; ds.
XX
OS Chimeric - Human immunodeficiency virus.
OS Chimeric - Simian virus.
XX
PN WO200176643-A1.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-US11372.
XX
PR 07-APR-2000; 2000US-195680P.
XX
PA (BAYU) BAYLOR COLLEGE MEDICINE.
XX
PI Orson FM, Kinsey BM, Bhogal BS;
XX
DR WPI; 2002-066308/09.

XX
PT Composition for oral delivery of vaccines, comprises expression vector
PT containing antigenic genomic sequence, bound to aggregated
PT protein-polycationic polymer conjugate or suspension -
XX
PS Disclosure; Page 76-79; 145pp; English.
XX
CC The invention relates to a composition comprising an expression vector
CC bound to an aggregated protein-polycationic polymer conjugate or
CC suspension. The expression vector contains a promoter polynucleotide
CC sequence operatively linked to a polynucleotide sequence encoding an
CC antigen which is a fragment of a gene or genome associated with an
CC infectious disease, cancer and autoimmune disease such as rheumatoid
CC arthritis, vasculitis, and multiple sclerosis, pathogenic genomes
CC consisting of bacterium, fungus, protozoa and virus such as human
CC immunodeficiency virus (HIV), herpes simplex virus (HSV), hepatitis C
CC virus (HCV), influenza and respiratory syncytial virus (RSV), and
CC optionally comprising a nucleotide sequence encoding a cytokine (or a
CC cytokine expression vector), is useful for inducing an immune response
CC (systemic and/or mucosal) in an organism. The cytokine expression vector
CC contains a sequence for granulocyte macrophage-colony stimulating factor
CC (GM-CSF) or interleukin-12 (IL-12). The polynucleotide sequences encoding
CC the antigen and the cytokine are under transcriptional control of same or
CC different promoter polynucleotide sequences. The expression vector, as a
CC DNA vaccine is useful for treating a condition in an organism. The
CC present sequence is simian-human immunodeficiency virus clone
CC SHIV-HXBcP 3.2 complete genomic DNA, related to the invention.
XX
SQ Sequence 10006 BP; 3459 A; 1824 C; 2505 G; 2218 T; 0 other;

Query Match 100.0%; Score 2564; DB 24; Length 10006;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2564; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2332 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 2391

Qy      61 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2392 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 2451

Qy     121 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2452 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 2511

Qy     181 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2512 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 2571

Qy     241 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2572 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 2631

Qy     301 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2632 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 2691

Qy     361 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2692 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 2751

Qy     421 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2752 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 2811

Qy     481 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2812 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 2871

Qy     541 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2872 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 2931

Qy     601 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2932 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 2991

Qy     661 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2992 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 3051

Qy     721 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3052 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 3111

Qy     781 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3112 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 3171

Qy     841 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3172 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 3231

Qy     901 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3232 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 3291

Qy     961 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3292 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 3351

Qy    1021 AAAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3352 AAAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACA 3411

Qy    1081 CATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATA 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3412 CATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATA 3471

Qy    1141 GTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3472 GTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAG 3531

Qy    1201 TGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCA 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3532 TGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCA 3591

Qy    1261 CCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTAT 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3592 CCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTAT 3651

Qy    1321 TATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGAT 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3652 TATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGAT 3711

Qy    1381 AGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAA 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3712 AGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAA 3771

Qy    1441 GCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAA 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3772 GCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAA 3831

Qy    1501 TATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAA 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3832 TATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAA 3891

Qy    1561 ATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAA 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3892 ATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAA 3951

Qy    1621 GGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTC 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3952 GGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTC 4011
```

```
Qy   1681 TTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4012 TTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAA 4071

Qy   1741 GAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGT 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4072 GAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGT 4131

Qy   1801 GATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACT 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4132 GATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACT 4191

Qy   1861 TGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCT 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4192 TGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCT 4251

Qy   1921 AGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTT 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4252 AGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTT 4311

Qy   1981 CTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAAC 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4312 CTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAAC 4371

Qy   2041 TTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGG 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4372 TTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGG 4431

Qy   2101 GTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAAT 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4432 GTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAAT 4491

Qy   2161 CAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTT 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4492 CAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTT 4551

Qy   2221 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTA 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4552 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTA 4611

Qy   2281 ATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTT 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4612 ATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTT 4671

Qy   2341 AAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAG 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4672 AAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAG 4731

Qy   2401 CTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTA 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4732 CTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTA 4791

Qy   2461 CCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGT 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4792 CCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGT 4851

Qy   2521 TCCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 2564
          |||||||||||||||||||||||||||||||||||||||||||
Db   4852 TCCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 4895
```

RESULT 2
AAV81865
ID AAV81865 standard; DNA; 10279 BP.
XX
AC AAV81865;
XX
DT 01-JUN-1999 (first entry)
XX
DE Nucleic acid sequence of SIVmac239 genome.
XX
KW Antigenic composition; primate; lentivirus; nef gene; vaccine;
KW infection; AIDS; SIVmac239; ss.
XX
OS Simian immunodeficiency virus.
XX
PN US5851813-A.
XX
PD 22-DEC-1998.
XX
PF 27-JAN-1994; 94US-0188583.
XX
PR 27-JAN-1994; 94US-0188583.
PR 12-JUL-1990; 90US-0551945.
PR 09-JUL-1991; 91US-0727494.
XX
PA (HARD) HARVARD COLLEGE.
XX
PI Desrosiers RC;
XX
DR WPI; 1999-080408/07.
DR P-PSDB; AAW89313, AAW89314, AAW89315, AAW89316, AAW89317, AAW89318,
DR AAW89319, AAW89320, AAW89321.
XX
PT Lentivirus antigenic compositions - containing lentivirus with nef
PT gene deletion
XX
PS Disclosure; Fig 1A-V; 93pp; English.
XX
CC The invention relates to an antigenic composition comprising an isolated
CC primate lentivirus whose genome contains an engineered non-revertible
CC null mutation in the nef gene, or an infectious DNA clone in a carrier.
CC The antigenic composition is used in vaccines against infection by the
CC lentivirus, e.g. AIDS. The present sequence represents the nucleic
CC acid sequence of the SIVmac 239 genome.
XX
SQ Sequence 10279 BP; 3466 A; 1935 C; 2570 G; 2308 T; 0 other;

Query Match 100.0%; Score 2564; DB 20; Length 10279;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2564; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2849 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 2908

Qy     61 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2909 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 2968

Qy    121 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2969 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 3028

Qy    181 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3029 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 3088

Qy    241 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3089 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 3148

Qy    301 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3149 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 3208

Qy    361 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3209 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 3268

Qy    421 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3269 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 3328

Qy    481 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   3329 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 3388

Qy    541 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3389 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 3448

Qy    601 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3449 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 3508

Qy    661 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3509 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 3568

Qy    721 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3569 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 3628

Qy    781 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3629 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 3688

Qy    841 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3689 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 3748

Qy    901 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3749 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 3808

Qy    961 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3809 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 3868

Qy   1021 AAAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3869 AAAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACA 3928

Qy   1081 CATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATA 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3929 CATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATA 3988

Qy   1141 GTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3989 GTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAG 4048

Qy   1201 TGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCA 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4049 TGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCA 4108

Qy   1261 CCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTAT 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4109 CCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTAT 4168

Qy   1321 TATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGAT 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4169 TATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGAT 4228

Qy   1381 AGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAA 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4229 AGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAA 4288

Qy   1441 GCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAA 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4289 GCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAA 4348

Qy   1501 TATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAA 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4349 TATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAA 4408

Qy   1561 ATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAA 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4409 ATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAA 4468
```

```
Qy   1621 GGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTC 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4469 GGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTC 4528

Qy   1681 TTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4529 TTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAA 4588

Qy   1741 GAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGT 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4589 GAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGT 4648

Qy   1801 GATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACT 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4649 GATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACT 4708

Qy   1861 TGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCT 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4709 TGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCT 4768

Qy   1921 AGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTT 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4769 AGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTT 4828

Qy   1981 CTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAAC 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4829 CTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAAC 4888

Qy   2041 TTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGG 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4889 TTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGG 4948

Qy   2101 GTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAAT 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4949 GTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAAT 5008

Qy   2161 CAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTT 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5009 CAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTT 5068

Qy   2221 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTA 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5069 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTA 5128

Qy   2281 ATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTT 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5129 ATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTT 5188

Qy   2341 AAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAG 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5189 AAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAG 5248

Qy   2401 CTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTA 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5249 CTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTA 5308

Qy   2461 CCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGT 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5309 CCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGT 5368

Qy   2521 TCCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 2564
          ||||||||||||||||||||||||||||||||||||||||||||
Db   5369 TCCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 5412
```

RESULT 3
AAQ24802
ID AAQ24802 standard; DNA; 10097 BP.
XX
AC AAQ24802;
XX
DT 06-JUL-1992 (first entry)

```
XX
DE   SIVmac239 nef-deletion.
XX
KW   Macaque; monkey; polymerase chain reaction;
KW   PCR; site-directed mutagenesis; retrovirus; null mutation; ss.
XX
OS   Simian immunodeficiency virus.
XX
FH   Key             Location/Qualifiers
FT   repeat_region   1..818
FT                   /*tag= a
FT                   /rpt_type= TERMINAL
FT                   /note= "i.e. 5' LTR"
FT   repeat_unit     1..517
FT                   /*tag= b
FT                   /rpt_type= OTHER
FT                   /note= "U3"
FT   repeat_unit     518..600
FT                   /*tag= c
FT                   /rpt_type= OTHER
FT                   /note= "R"
FT   repeat_unit     601..818
FT                   /*tag= d
FT                   /rpt_type= OTHER
FT                   /note= "U5"
FT   primer_bind     822..849
FT                   /*tag= e
FT                   /standard_name= tRNA_PBS
FT   CDS             1053..2585
FT                   /*tag= f
FT                   /product= gag
FT   CDS             2228..5410
FT                   /*tag= g
FT                   /product= pol
FT   CDS             5340..5984
FT                   /*tag= h
FT                   /product= vif
FT   CDS             5812..6150
FT                   /*tag= i
FT                   /product= vpx
FT   CDS             6051..6456
FT                   /*tag= j
FT                   /product= vpr
FT   exon            6302..6597
FT                   /*tag= k
FT                   /product= tat
FT                   /note= "full-length product obtained by splicing"
FT   exon            6528..6597
FT                   /*tag= l
FT                   /product= rev
FT                   /note= "full-length product obtained by splicing"
FT   CDS             6604..9243
FT                   /*tag= m
FT                   /product= env
FT   exon            8803..8902
FT                   /*tag= n
FT                   /product= tat
FT                   /note= "see above"
FT   exon            8803..9059
FT                   /*tag= o
FT                   /product= rev
FT                   /note= "see above"
FT   CDS             9077..9686
FT                   /*tag= p
FT                   /product= nef
FT   repeat_region   9280..10097
FT                   /*tag= q
FT                   /rpt_type= TERMINAL
FT                   /note= "i.e. 3' LTR"
FT   misc_signal     415..424
FT                   /*tag= r
FT                   /standard_name= NF_Kappa_B
FT   GC_signal       429..438
FT                   /*tag= s
FT                   /standard_name= Sp1_binding_site
FT   GC_signal       440..449
FT                   /*tag= t
FT                   /standard_name= Sp1_binding_site
FT   GC_signal       451..460
FT                   /*tag= u
FT                   /standard_name= Sp1_binding_site
FT   GC_signal       462..471
FT                   /*tag= v
FT                   /standard_name= Sp1_binding_site
FT   TATA_signal     488..494
FT                   /*tag= w
FT   polyA_signal    9950..9955
FT                   /*tag= x
XX
PN   WO9200987-A.
XX
PD   23-JAN-1992.
XX
PF   10-JUL-1991;   91WO-US04884.
XX
PR   12-JUL-1990;   90US-0551945.
XX
PA   (HARD ) HARVARD COLLEGE.
XX
PI   Desrosiers RC;
XX
DR   WPI; 1992-056816/07.
DR   P-PSDB; AAR22365-R22371, AAR24126-7.
XX
PT   Primate lentivirus vaccine protecting against AIDS - and primate
PT   lentiviruses and their DNA clones contg. null mutations, useful for
PT   producing vaccine
XX
PS   Claim 1; Fig 1; 51pp; English.
XX
CC   Parental virus SIVmac239 was isolated from a macaque monkey (see
CC   AAQ22487). An SstI fragment of p239SpE3' contg. the C-terminus of
CC   the nef gene was cloned into M13 and subjected to site-directed
CC   mutagenesis. A 73mer primer was used which was complementary to
CC   bases 9215 through 9250 and 9433 through 9469. Since bases 9251
CC   through 9432 were not included in the primer, their complement was
CC   not included in the newly synthesised negative strand. Successful
CC   deletion was confirmed by DNA sequencing. The fragment contg. the
CC   deletion was cloned into p239SpE3' to yield p239S-E3'(nef-deletion)
CC   which was digested with SphI and ligated to SphI-cut p239SpSp5'.
CC   The ligation product was used to directly transfect cultured cells.
CC   See also AAQ21075-8.
XX
SQ   Sequence 10097 BP; 3387 A; 1911 C; 2527 G; 2272 T; 0 other;

  Query Match             99.9%;  Score 2562.4;  DB 13;  Length 10097;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2563;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

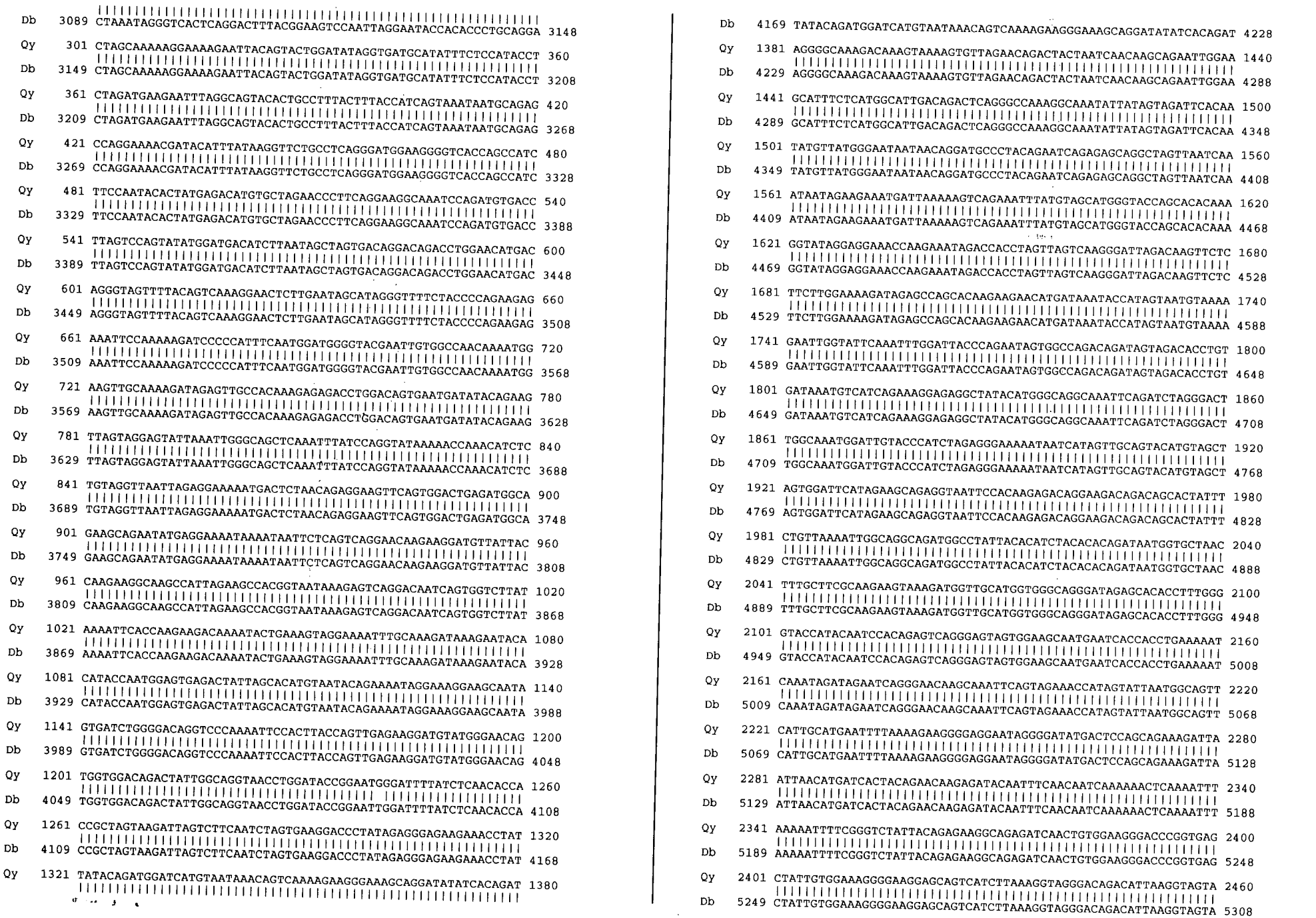
Qy      1 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2849 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 2908

Qy     61 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2909 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 2968

Qy    121 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2969 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 3028

Qy    181 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3029 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 3088

Qy    241 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 300
```


Db 3089 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 3148
Qy 301 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 360
Db 3149 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 3208
Qy 361 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 420
Db 3209 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 3268
Qy 421 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 480
Db 3269 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 3328
Qy 481 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 540
Db 3329 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 3388
Qy 541 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 600
Db 3389 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 3448
Qy 601 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 660
Db 3449 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 3508
Qy 661 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 720
Db 3509 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 3568
Qy 721 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 780
Db 3569 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 3628
Qy 781 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 840
Db 3629 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 3688
Qy 841 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 900
Db 3689 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 3748
Qy 901 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 960
Db 3749 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 3808
Qy 961 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 1020
Db 3809 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 3868
Qy 1021 AAAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACA 1080
Db 3869 AAAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACA 3928
Qy 1081 CATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATA 1140
Db 3929 CATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATA 3988
Qy 1141 GTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAG 1200
Db 3989 GTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAG 4048
Qy 1201 TGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCA 1260
Db 4049 TGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATTGGATTTTATCTCAACACCA 4108
Qy 1261 CCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTAT 1320
Db 4109 CCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTAT 4168
Qy 1321 TATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGAT 1380
Db 4169 TATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGAT 4228
Qy 1381 AGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAA 1440
Db 4229 AGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAA 4288
Qy 1441 GCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAA 1500
Db 4289 GCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAA 4348
Qy 1501 TATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAA 1560
Db 4349 TATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAA 4408
Qy 1561 ATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAA 1620
Db 4409 ATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAA 4468
Qy 1621 GGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTC 1680
Db 4469 GGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTC 4528
Qy 1681 TTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAA 1740
Db 4529 TTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAA 4588
Qy 1741 GAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGT 1800
Db 4589 GAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGT 4648
Qy 1801 GATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACT 1860
Db 4649 GATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACT 4708
Qy 1861 TGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCT 1920
Db 4709 TGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCT 4768
Qy 1921 AGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTT 1980
Db 4769 AGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTT 4828
Qy 1981 CTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAAC 2040
Db 4829 CTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAAC 4888
Qy 2041 TTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGG 2100
Db 4889 TTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGG 4948
Qy 2101 GTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAAT 2160
Db 4949 GTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAAT 5008
Qy 2161 CAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTT 2220
Db 5009 CAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTT 5068
Qy 2221 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTA 2280
Db 5069 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTA 5128
Qy 2281 ATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTT 2340
Db 5129 ATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTT 5188
Qy 2341 AAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAG 2400
Db 5189 AAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAG 5248
Qy 2401 CTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTA 2460
Db 5249 CTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTA 5308

```
Qy   2461 CCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGT 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5309 CCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGT 5368

Qy   2521 TCCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 2564
          |||||||||||||||||||||||||||||||||||||||||||
Db   5369 TCCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 5412


RESULT 4
AAQ22487
ID   AAQ22487 standard; DNA; 10279 BP.
XX
AC   AAQ22487;
XX
DT   06-JUL-1992  (first entry)
XX
DE   SIVmac239 proviral genome.
XX
KW   Macaque; monkey; mac239; polymerase chain reaction;
KW   PCR; site-directed mutagenesis; retrovirus; ss.
XX
OS   Simian immunodeficiency virus.
XX
FH   Key             Location/Qualifiers
FT   repeat_region   1..818
FT                   /*tag= a
FT                   /rpt_type= TERMINAL
FT                   /note= "i.e. 5' LTR"
FT   repeat_unit     1..517
FT                   /*tag= b
FT                   /rpt_type= OTHER
FT                   /note= "U3"
FT   repeat_unit     518..600
FT                   /*tag= c
FT                   /rpt_type= OTHER
FT                   /note= "R"
FT   repeat_unit     601..818
FT                   /*tag= d
FT                   /rpt_type= OTHER
FT                   /note= "U5"
FT   primer_bind     822..849
FT                   /*tag= e
FT                   /standard_name= tRNA_PBS
FT   CDS             1053..2585
FT                   /*tag= f
FT                   /product= gag
FT   CDS             2228..5410
FT                   /*tag= g
FT                   /product= pol
FT   CDS             5340..5984
FT                   /*tag= h
FT                   /product= vif
FT   CDS             5812..6150
FT                   /*tag= i
FT                   /product= vpx
FT   CDS             6051..6456
FT                   /*tag= j
FT                   /product= vpr
FT   exon            6302..6597
FT                   /*tag= k
FT                   /product= tat
FT                   /note= "full-length product obtained by splicing"
FT   exon            6528..6597
FT                   /*tag= l
FT                   /product= rev
FT                   /note= "full-length product obtained by splicing"
FT   CDS             6604..9243
FT                   /*tag= m
FT                   /product= env
FT   exon            8803..8902
FT                   /*tag= n
FT                   /product= tat
FT                   /note= "see above"
FT   exon            8803..9059
FT                   /*tag= o
FT                   /product= rev
FT                   /note= "see above"
FT   CDS             9077..9868
FT                   /*tag= p
FT                   /product= nef
FT   repeat_region   9462..10279
FT                   /*tag= q
FT                   /rpt_type= TERMINAL
FT                   /note= "i.e. 3' LTR"
FT   misc_signal     415..424
FT                   /*tag= r
FT                   /standard_name= NF_Kappa_B
FT   GC_signal       429..438
FT                   /*tag= s
FT                   /standard_name= Sp1_binding_site
FT   GC_signal       440..449
FT                   /*tag= t
FT                   /standard_name= Sp1_binding_site
FT   GC_signal       451..460
FT                   /*tag= u
FT                   /standard_name= Sp1_binding_site
FT   GC_signal       462..471
FT                   /*tag= v
FT                   /standard_name= Sp1_binding_site
FT   TATA_signal     488..494
FT                   /*tag= w
FT   polyA_signal    10132..10137
FT                   /*tag= x
XX
PN   WO9200987-A.
XX
PD   23-JAN-1992.
XX
PF   10-JUL-1991;   91WO-US04884.
XX
PR   12-JUL-1990;   90US-0551945.
XX
PA   (HARD ) HARVARD COLLEGE.
XX
PI   Desrosiers RC;
XX
DR   WPI; 1992-056816/07.
DR   P-PSDB; AAR22365-R22372, AAR24126.
XX
PT   Primate lentivirus vaccine protecting against AIDS - and primate
PT   lentiviruses and their DNA clones contg. null mutations, useful for
PT   producing vaccine
XX
PS   Disclosure; Fig 1; 51pp; English.
XX
CC   Cell-free serum samples from a macaque monkey exhibiting symptoms
CC   characteristic of SIV infection were co-cultivated with HuT-78
CC   cells. Infectious SIVmac239 virus was identified in the cell
CC   supernatant. Total cell DNA was prepared from SIVmac239-infected
CC   cells and digested with EcoRI. An EMBL-4 library was constructed
CC   from 10-20kb EcoRI fragments (EcoRI is a non-cutter of SIVmac239).
CC   The library was screened with pK2 BamA as probe and a full-length
CC   molecular clone was isolated and sequenced. Then, EMBL-SIVmac239
CC   was digested with SphI and a 6706bp fragment, contg. the SphI site
CC   in the left flanking cellular DNA sequence to viral nucleotide no.
CC   6451, was inserted in vector pBS(+) to produce subclone p239SpSp5'.
CC   In a separate reaction, EMBL-SIVmac239 was digested with EcoRI and
CC   SphI. A 6361bp fragment from viral nucleotide 6452 to the EcoRI
CC   site in the right flanking cellular sequence, was inserted in
CC   pBS(-) to produce subclone p239SpE3'. These subclones were used to
CC   generate the full-length genomic sequence and to produce the
CC   preferred null-mutations of the invention. See Q24802 for
CC   nef-deletion mutant. See also Q21075-8.
XX
```

```
SQ   Sequence 10279 BP; 3465 A; 1936 C; 2569 G; 2309 T; 0 other;

  Query Match           99.9%;  Score 2562.4;  DB 13;  Length 10279;
  Best Local Similarity 100.0%;  Pred. No. 0;
  Matches 2563;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2849 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 2908

Qy      61 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2909 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 2968

Qy     121 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2969 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 3028

Qy     181 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3029 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 3088

Qy     241 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3089 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 3148

Qy     301 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3149 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 3208

Qy     361 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3209 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 3268

Qy     421 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3269 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 3328

Qy     481 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3329 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 3388

Qy     541 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3389 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 3448

Qy     601 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3449 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 3508

Qy     661 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3509 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 3568

Qy     721 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3569 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 3628

Qy     781 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3629 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 3688

Qy     841 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3689 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 3748

Qy     901 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3749 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 3808

Qy     961 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3809 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 3868

Qy    1021 AAAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3869 AAAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACA 3928

Qy    1081 CATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATA 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3929 CATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATA 3988

Qy    1141 GTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3989 GTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAG 4048

Qy    1201 TGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCA 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4049 TGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATTGGATTTTATCTCAACACCA 4108

Qy    1261 CCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTAT 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4109 CCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTAT 4168

Qy    1321 TATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGAT 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4169 TATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGAT 4228

Qy    1381 AGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAA 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4229 AGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAA 4288

Qy    1441 GCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAA 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4289 GCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAA 4348

Qy    1501 TATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAA 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4349 TATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAA 4408

Qy    1561 ATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAA 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4409 ATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAA 4468

Qy    1621 GGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTC 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4469 GGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTC 4528

Qy    1681 TTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4529 TTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAA 4588

Qy    1741 GAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGT 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4589 GAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGT 4648

Qy    1801 GATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACT 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4649 GATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACT 4708

Qy    1861 TGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCT 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4709 TGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCT 4768

Qy    1921 AGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTT 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4769 AGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTT 4828

Qy    1981 CTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAAC 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4829 CTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAAC 4888

Qy    2041 TTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGG 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4889 TTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGG 4948
```

```
Qy   2101 GTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAAT 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4949 GTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAAT 5008

Qy   2161 CAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTT 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5009 CAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTT 5068

Qy   2221 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTA 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5069 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTA 5128

Qy   2281 ATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTT 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5129 ATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTT 5188

Qy   2341 AAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAG 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5189 AAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAG 5248

Qy   2401 CTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTA 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5249 CTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTA 5308

Qy   2461 CCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGT 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5309 CCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGT 5368

Qy   2521 TCCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 2564
          |||||||||||||||||||||||||||||||||||||||||||
Db   5369 TCCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 5412
```

```
RESULT 5
AAD25516
ID   AAD25516 standard; DNA; 10274 BP.
XX
AC   AAD25516;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Simian-Human immunodeficiency virus (SHIV) clone 1A11 complete genome.
XX
KW   HIV; human immunodeficiency virus; simian virus; cytostatic; cancer;
KW   immunosuppressive; virucide; antibacterial; fungicide; protozoacide;
KW   antirheumatic; antiinflammatory; antiarthritic; rheumatoid arthritis;
KW   neuroprotective; multiple sclerosis; immune response; vasotropic;
KW   vaccine; gene therapy; autoimmune disease; vasculitis; ds.
XX
OS   Chimeric - Human immunodeficiency virus.
OS   Chimeric - Simian virus.
XX
PN   WO200176643-A1.
XX
PD   18-OCT-2001.
XX
PF   06-APR-2001; 2001WO-US11372.
XX
PR   07-APR-2000; 2000US-195680P.
XX
PA   (BAYU ) BAYLOR COLLEGE MEDICINE.
XX
PI   Orson FM,  Kinsey BM,  Bhogal BS;
XX
DR   WPI; 2002-066308/09.
XX
PT   Composition for oral delivery of vaccines, comprises expression vector
PT   containing antigenic genomic sequence, bound to aggregated
PT   protein-polycationic polymer conjugate or suspension  -
XX
PS   Disclosure; Page 82-84; 145pp; English.
XX
CC   The invention relates to a composition comprising an expression vector
CC   bound to an aggregated protein-polycationic polymer conjugate or
CC   suspension. The expression vector contains a promoter polynucleotide
CC   sequence operatively linked to a polynucleotide sequence encoding an
CC   antigen which is a fragment of a gene or genome associated with an
CC   infectious disease, cancer and autoimmune disease such as rheumatoid
CC   arthritis, vasculitis, and multiple sclerosis, pathogenic genomes
CC   consisting of bacterium, fungus, protozoa and virus such as human
CC   immunodeficiency virus (HIV), herpes simplex virus (HSV), hepatitis C
CC   virus (HCV), influenza and respiratory syncytial virus (RSV), and
CC   optionally comprising a nucleotide sequence encoding a cytokine (or a
CC   cytokine expression vector), is useful for inducing an immune response
CC   (systemic and/or mucosal) in an organism. The cytokine expression vector
CC   contains a sequence for granulocyte macrophage-colony stimulating factor
CC   (GM-CSF) or interleukin-12 (IL-12). The polynucleotide sequences encoding
CC   the antigen and the cytokine are under transcriptional control of same or
CC   different promoter polynucleotide sequences. The expression vector, as a
CC   DNA vaccine is useful for treating a condition in an organism. The
CC   present sequence is simian-human immunodeficiency virus (SHIV) clone
CC   1A11 complete genomic DNA, related to the invention.
XX
SQ   Sequence 10274 BP; 3475 A; 1931 C; 2559 G; 2309 T; 0 other;

  Query Match             97.9%;  Score 2510.2;  DB 24;  Length 10274;
  Best Local Similarity   98.7%;  Pred. No. 0;
  Matches 2530;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;
```

```
Qy      2 TTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCAA 61
          |||||||||||||||||||||||||||||||||||||||||||||||||||| | ||||||||
Db   2849 TTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAAGTTGGACCAA 2908

Qy     62 AATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAAA 121
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2909 AATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAAA 2968

Qy    122 AGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCCA 181
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2969 AGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCCA 3028

Qy    182 CATTTGCTATAAAGAAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAAC 241
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   3029 CATTTGCTATAAAGAAAAAAGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAAC 3088

Qy    242 TAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGAC 301
          ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   3089 TAAATAGGGTCACTCAGGACTTTACAGAAGTCCAATTAGGAATACCACACCCTGCAGGAC 3148

Qy    302 TAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCTC 361
          ||||||||||||||| |||||||||||||||||||| |||||||||||||||||||||||||
Db   3149 TAGCAAAAAGGAAAAGGATTACAGTACTGGATATAGGTGACGCATATTTCTCCATACCTC 3208

Qy    362 TAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAGC 421
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3209 TAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAGC 3268

Qy    422 CAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATCT 481
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   3269 CAGGAAAACGATACATTTATAAGGTTCTGCCTCAAGGATGGAAGGGGTCACCAGCCATCT 3328

Qy    482 TCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACCT 541
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3329 TCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACCT 3388

Qy    542 TAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGACA 601
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   3389 TAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAAACCTGGAACATGACA 3448

Qy    602 GGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAGA 661
          |||||||||||||| |||||||||||| ||||||||||||||||||||||||||||||||||
Db   3449 GGGTAGTTTTACAGTTAAAGGAACTCTTAAATAGCATAGGGTTTTCTACCCCAGAAGAGA 3508

Qy    662 AATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGGA 721
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3509 AATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGGA 3568
```

```
Qy    722 AGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAGT 781
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3569 AGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAGT 3628

Qy    782 TAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTCT 841
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3629 TAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTCT 3688

Qy    842 GTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCAG 901
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3689 GTAGGCTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCAG 3748

Qy    902 AAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTACC 961
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3749 AAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTACC 3808

Qy    962 AAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTATA 1021
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   3809 AAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAATCAGGACAATCAGTGGTCTTATA 3868

Qy   1022 AAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACAC 1081
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3869 AAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACAC 3928

Qy   1082 ATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATAG 1141
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   3929 ATACCAATGGAGTTAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATAG 3988

Qy   1142 TGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAGT 1201
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3989 TGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAGT 4048

Qy   1202 GGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCAC 1261
          |||||||||||||||||||||||||||||||| |||||| ||||||||||||||||||||
Db   4049 GGTGGACAGACTATTGGCAGGTAACCTGGATACCAGAATGGAATTTTATCTCAACACCAC 4108

Qy   1262 CGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTATT 1321
          | |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4109 CACTAGTGAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTATT 4168

Qy   1322 ATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGATA 1381
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4169 ATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGATA 4228

Qy   1382 GGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAAG 1441
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   4229 GGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAGGCAGAATTGGAAG 4288

Qy   1442 CATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAAT 1501
          |||| |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   4289 CATTCCTCATGGCATTAACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAAT 4348

Qy   1502 ATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAAA 1561
          |||||||||||||||||||||||||||||||||||||||||||||| |||||||| ||||
Db   4349 ATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGACTAGTTAACCAAA 4408

Qy   1562 TAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAAG 1621
          ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4409 TAATAAAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAAG 4468

Qy   1622 GTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTCT 1681
          |||||||||||||||||| |||||||||||||||||| ||||||||||||||||||||||||
Db   4469 GTATAGGAGGAAACCAAGAAGTAGACCACCTAGTTAGTCAGGGGATTAGACAAGTTCTCT 4528

Qy   1682 TCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAAG 1741
          |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db   4529 TCTTGGAAAAGATAGAGCCAGCACAAGAAGAGCATGATAAATACCATAGTAATGTAAAAG 4588

Qy   1742 AATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGTG 1801
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4589 AATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGTG 4648

Qy   1802 ATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACTT 1861
          |||||||||||||||||||||| |||||||||||||||| ||||||||||||||||||||
Db   4649 ATAAATGTCATCAGAAAGGAGAAGCTATACATGGGCAGGTAAATTCAGATCTAGGGACTT 4708

Qy   1862 GGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCTA 1921
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4709 GGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCTA 4768

Qy   1922 GTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTTC 1981
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4769 GTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTTC 4828

Qy   1982 TGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAACT 2041
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4829 TGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAACT 4888

Qy   2042 TTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGGG 2101
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4889 TTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGGG 4948

Qy   2102 TACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAATC 2161
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4949 TACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAATC 5008

Qy   2162 AAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTTC 2221
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5009 AAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTTC 5068

Qy   2222 ATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTAA 2281
          |||  ||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db   5069 ATTCGATGAATTTTAAAAGAAGGGGAGGAACAGGGGATATGACTCCAGCAGAAAGATTAA 5128

Qy   2282 TTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTTA 2341
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   5129 TTAACATGATCACTACAGAACAAGAAATACAATTTCAACAATCAAAAAACTCAAAATTTA 5188

Qy   2342 AAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAGC 2401
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5189 AAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAGC 5248

Qy   2402 TATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTAC 2461
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5249 TATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTAC 5308

Qy   2462 CCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGTT 2521
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   5309 CCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGATGGATAGCAGTT 5368

Qy   2522 CCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 2564
          |||||||||||||||||||||||||||||||||||||||||||
Db   5369 CCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 5411
```

RESULT 6
AAN80860
ID AAN80860 standard; cDNA; 9601 BP.
XX
AC AAN80860;
XX
DT 15-NOV-1990 (first entry)
XX
DE Sequence corresp. to the genome of SIV isolate STLV-III Mac 142-83
DE in clone SIV-1.
XX
KW pSIV-1.1; pSIV-1.2; STLV-III Mac 142-83; ss.
XX
OS Simian immunodeficiency virus.
XX

FH	Key	Location/Qualifiers
FT	CDS	551..2068
FT		/*tag= a
FT		/gene="gagd"
FT	CDS	1726..4893

```
FT                   /*tag=  b
FT                   /gene="pol"
FT   CDS             4826..5467
FT                   /*tag=  c
FT                   /gene="Q"
FT   CDS             5298..5633
FT                   /*tag=  d
FT                   /gene="X"
FT   CDS             5637..5939
FT                   /*tag=  e
FT                   /gene="R"
FT   CDS             5788..6084
FT                   /*tag=  f
FT                   /gene="tat1"
FT   CDS             6014..6130
FT                   /*tag=  g
FT                   /gene="art1"
FT   CDS             6090..8297
FT                   /*tag=  h
FT                   /gene="env"
FT   CDS             8298..8732
FT                   /*tag=  i
FT                   /gene="env ctd."
FT   CDS             8296..8391
FT                   /*tag=  j
FT                   /gene="tat2"
FT   CDS             8294..8548
FT                   /*tag=  k
FT                   /gene="art2"
FT   CDS             8569..9354
FT                   /*tag=  l
FT                   /gene="F"
XX
PN   WO8805440-A.
XX
PD   28-JUL-1988.
XX
PF   15-JAN-1988;   88WO-FR00025.
XX
PR   15-APR-1987;   87FR-0005398.
XX
PA   (INSP ) INST PASTEUR.
PA   (ALIZ/) ALIZON M.
XX
PI   Alizon M,  Montagnier L,  Guetard D,  Clavfl F,  Sonigo P,  Guyader M;
PI   Tiollais P,  Chakrabarti L,  Desrosiers R;
XX
DR   WPI; 1988-220290/31.
DR   P-PSDB; AAP80802, AAP81783, AAP81784, AAP81785, AAP81691, AAP81756,
DR   AAP81757, AAP81758, AAP81759, AAP81760, AAP81761, AAP81762.
XX
PT   New peptide(s) with immunological properties of HIV-2 envelope protein -
PT   having the structure of simian immune deficiency virus proteins,
PT   useful in diagnosis and of vaccine components
XX
PS   Claim 23; Fig 1B; 86pp; French.
XX
CC   AAN80860 is contained in lambda SIV-1 which was deposited on 15/4/87 at
CC   the CNCM under numbers I-658 (pSIV-1.1) and I-659 (pSIV-1.2). It is the
CC   cDNA to SIV-1 genomic RNA. Recombinant DNA contg. all or part of the
CC   cDNA from this sequence inserted into a vector are claimed. Labelled
CC   fragments of cDNA can be used as hybridisation probes. New peptides which
CC   have immunological properties in common with those of the peptide
CC   skeleton of the envelope protein of HIV-2 and also have a peptide
CC   structure in common with that of SIV-1 glycoprotein are claimed.
CC   Antigenic and immunogenic conjugates contg. the peptides and a kit to
CC   detect HIV-2 in biological fluids are new. The peptides are useful for in
CC   vitro diagnosis of HIV-2 infection and some of them can be used as
CC   components of immunogens and vaccines against HIV. Antibodies raised
CC   against them can be used for treatment of AIDS.
XX
SQ   Sequence 9601 BP; 3277 A; 1809 C; 2378 G; 2136 T; 1 other;
```

```
  Query Match             96.2%;  Score 2467;  DB 9;  Length 9601;
  Best Local Similarity   97.7%;  Pred. No. 0;
  Matches 2503;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy       2 TTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCAA 61
          ||||||||||||| ||||||||||||||     ||||||||||||||||||||||||||||
Db    2337 TTCCCATAGCTAAGGTAGAGCCTGTAAAGTCGCCCTTAAAGCCAGGAAAGGATGGACCAA 2396

Qy      62 AATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAAA 121
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2397 AATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAAA 2456

Qy     122 AGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCA 181
          |||||||||| |||||||||||||||||||||||||||||||||||||| ||||||||||
Db    2457 AGATGGAAAAAGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATATAACACCCCA 2516

Qy     182 CATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAAC 241
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db    2517 CATTTGCTATAAAGAAAAAGGATAAAAACAAATGGAGAATGCTGATAGATTTTAGGGAAC 2576

Qy     242 TAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGAC 301
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    2577 TAAATAGGGTCACTCAAGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGAC 2636

Qy     302 TAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCTC 361
          |||||||||||||||| |||||||||||||||||||||| |||||||||| |||||||
Db    2637 TAGCAAAAAGGAAAAGGATTACAGTACTGGATATAGGTGACGCATATTTCTCTATACCTC 2696

Qy     362 TAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAGC 421
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2697 TAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAGC 2756

Qy     422 CAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATCT 481
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2757 CAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATCT 2816

Qy     482 TCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACCT 541
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2817 TCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACCT 2876

Qy     542 TAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGACA 601
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2877 TAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGACA 2936

Qy     602 GGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAGA 661
          |||||||||||||| ||| |||||||| |||||||||||||||  ||||||||||||
Db    2937 GGGTAGTTTTACAGTTAAAAGAACTCTTAAATAGCATAGGGTTTTCATCCCCAGAAGAGA 2996

Qy     662 AATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGGA 721
          |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db    2997 AATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCGACAAAATGGA 3056

Qy     722 AGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAGT 781
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db    3057 AGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAACT 3116

Qy     782 TAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTCT 841
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3117 TACTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTCT 3176

Qy     842 GTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCAG 901
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3177 GTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCAG 3236

Qy     902 AAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTACC 961
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3237 AAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTACC 3296

Qy     962 AAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTATA 1021
          ||||| |||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db    3297 AAGAAAGCAAGCCATTAGAAGCCACGGTGATAAAGAGTCAGGACAATCAGTGGTCTTATA 3356

Qy    1022 AAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACAC 1081
```

```
Db   3357 AAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACAC 3416
Qy   1082 ATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATAG 1141
Db   3417 ATACCAATGGAGTTAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATAG 3476
Qy   1142 TGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAGT 1201
Db   3477 TGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAGT 3536
Qy   1202 GGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCAC 1261
Db   3537 GGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTCATCTCAACACCAC 3596
Qy   1262 CGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTATT 1321
Db   3597 CATTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTATT 3656
Qy   1322 ATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGATA 1381
Db   3657 ATGTAGATGGATCATGTAGTAAACAGTCAAAAGAAGGAAAAGCAGGATATATCACAGACA 3716
Qy   1382 GGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAAG 1441
Db   3717 GGGGCAAAGACAAGGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAAG 3776
Qy   1442 CATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAAT 1501
Db   3777 CATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGACTCACAAT 3836
Qy   1502 ATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAAA 1561
Db   3837 ATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAACCAAA 3896
Qy   1562 TAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAAG 1621
Db   3897 TAATAGAAGAAATGATCAAAAAGACAGAAATTTATGTGGCATGGGTACCAGCACACAAAG 3956
Qy   1622 GTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTCT 1681
Db   3957 GTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTCT 4016
Qy   1682 TCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAAG 1741
Db   4017 TCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATAGTAAATACCATAGTAACATAAAAG 4076
Qy   1742 AATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGTG 1801
Db   4077 AATTGGTATTCAAATTTGGATTACCCAGACTAGTGGCCAAACAGATAGTAGACACATGTG 4136
Qy   1802 ATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACTT 1861
Db   4137 ATAAATGTCATCAAAAAGGAGAAGCTATACATGGGCAGGTAAATTCAGACCTAGGGACTT 4196
Qy   1862 GGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCTA 1921
Db   4197 GGCAAATGGATTGTACCCATCTAGAGGGAAAAATAGTCATAGTTGCAGTACATGTAGCTA 4256
Qy   1922 GTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTTC 1981
Db   4257 GTGGATTCATAGAAGCAGAAGTAATTCCACAAGAAACAGGAAGACAGACAGCACTATTTC 4316
Qy   1982 TGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAACT 2041
Db   4317 TGTTAAAATTGGCAAGCAGATGGCCTATTACACATCTGCACACAGATAATGGTGCTAACT 4376
Qy   2042 TTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGGG 2101
Db   4377 TTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGGG 4436
Qy   2102 TACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAATC 2161
Db   4437 TACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAATC 4496
Qy   2162 AAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTTC 2221
Db   4497 AAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTTC 4556
Qy   2222 ATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTAA 2281
Db   4557 ATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTAA 4616
Qy   2282 TTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTTA 2341
Db   4617 TTAACATGATCACTACAGAACAAGAAATACAATTTCAACAATCAAAAAACTCAAAATTTA 4676
Qy   2342 AAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAGC 2401
Db   4677 AAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAGCTGTGGAAGGGACCCGGTGAGC 4736
Qy   2402 TATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTAC 2461
Db   4737 TATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGAACAGACATTAAGGTAGTAC 4796
Qy   2462 CCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGTT 2521
Db   4797 CCAGGAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGATGGATAGTAGTT 4856
Qy   2522 CCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 2564
Db   4857 CCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 4899
```

```
RESULT 7
AAT61085
ID   AAT61085 standard; DNA; 10163 BP.
XX
AC   AAT61085;
XX
DT   15-OCT-1997  (first entry)
XX
DE   Full-length HIV-2KR proviral DNA.
XX
KW   HIV-2; HIV-2KR; provirus; gene therapy; vector; vaccine; AIDS; ss.
XX
OS   Human immunodeficiency virus type 2.
XX
FH   Key             Location/Qualifiers
FT   provirus        1..10163
FT                   /*tag=  a
FT                   /note= "HIV-2KR"
FT   LTR             1..1070
FT                   /*tag=  b
FT                   /note= "5'LTR (Claim 14)"
FT   CDS             1071..2636
FT                   /*tag=  c
FT                   /note= "gag gene"
FT   CDS             2294..6038
FT                   /*tag=  d
FT                   /note= "pol gene"
FT   CDS             5391..6038
FT                   /*tag=  e
FT                   /note= "vif gene"
FT   CDS             6201..6518
FT                   /*tag=  f
FT                   /note= "vpr gene"
FT   CDS             6364..6650
FT                   /*tag=  g
FT                   /note= "tat gene 5' region"
FT   intron          6651..8795
FT                   /*tag=  h
FT                   /note= "intron interrupts tat and rev genes"
FT   CDS             8796..8892
FT                   /*tag=  i
FT                   /note= "tat gene 3' region"
```

```
FT   CDS             6590..6650
FT                   /*tag=  j
FT                   /note= "rev gene 5' region"
FT   CDS             8796..9265
FT                   /*tag=  k
FT                   /note= "rev gene 3' region"
FT   CDS             6660..9233
FT                   /*tag=  l
FT                   /note= "env gene"
FT   CDS             9067..9828
FT                   /*tag=  m
FT                   /note= "nef gene"
XX
PN   WO9705242-A1.
XX
PD   13-FEB-1997.
XX
PF   09-JUL-1996;    96WO-US11445.
XX
PR   07-JUN-1996;    96US-0659251.
PR   26-JUL-1995;    95US-0001441.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Kraus G,  Poeschla E,  Talbott R,  Wong-Staal F;
XX
DR   WPI; 1997-145681/13.
DR   P-PSDB; AAW13052-59.
XX
PT   Human immunodeficiency virus type II proviral clone - useful to
PT   develop prods. for study, diagnosis, prevention and treatment of HIV
PT   infections
XX
PS   Claim 1; Page 79-84; 128pp; English.
XX
CC   An isolated HIV-2 provirus (T61085), designated HIV-2KR and
CC   deposited as GenBank No. U22047, comprises a full-length HIV-2
CC   genome.  It was obtd. from the genomic DNA of Molt 4 Clone 8
CC   lymphoblastic cells infected with HIV-2PEI2, a strain isolated from
CC   the peripheral blood mononuclear cells of an HIV-2 seropositive
CC   AIDS patient.  Sequencing showed that the rev gene extends for
CC   nearly 69 amino acids past the end of most other HIV-2/SIV rev
CC   genes, and that the 5'LTR contains a deletion of 9-10 bp that is
CC   not present in other HIV-2/SIV isolates.  HIV-2KR is capable of
CC   infecting macaque peripheral blood lymphocytes in vitro and
CC   produces productive and persistent, albeit attenuated, infection in
CC   Macacca nemestrina.  Proviral constructs are useful for the
CC   generation of in vitro diagnostic reagents, gene therapy vectors
CC   and vaccines, and for the generation of HIV-2 based packaging cell
CC   lines.
XX
SQ   Sequence 10163 BP; 3456 A; 2046 C; 2540 G; 2121 T; 0 other;

  Query Match             68.3%;  Score 1751.8;  DB 18;  Length 10163;
  Best Local Similarity   80.2%;  Pred. No. 0;
  Matches 2056;  Conservative    0;  Mismatches  507;  Indels    0;  Gaps    0;

Qy       2 TTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCAA 61
Db    2901 TACCAGTCGCCAAGGTAGACCCGATAAAAGTAATACTGAAACCAGGAAAAGATGGACCAA 2960

Qy      62 AATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAAA 121
Db    2961 AAGTAAGACAATGGCCTCTAACAAAAGAAAAGATAGAGGCACTAAAAGAAATCTGTGAAA 3020

Qy     122 AGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCCA 181
Db    3021 AAATGGAAAGAGAAGGCCAGCTAGAGGAAGCTCCCCCAACTAATCCTTATAATACCCCCA 3080

Qy     182 CATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAAC 241
Db    3081 CATTTGCAATTAAGAAAAAGGACAAAAACAAATGGAGAATGCTAATAGATTTTAGAGAAC 3140

Qy     242 TAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGAC 301
Db    3141 TAAATAAGGTAACTCAAGAGTTCACAGAAATTCAGTTAGGAATTCCACACCCAGCAGGAT 3200

Qy     302 TAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCTC 361
Db    3201 TAGCCAAGAAAAGAAGAATTACTGTACTAGATATAGGGGATGCCTACTTTTCCATACCAC 3260

Qy     362 TAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAGC 421
Db    3261 TACATGAGGACTTTAGACAATATACTGCATTTACTCTACCAACAGTGAACAATGCAGAAC 3320

Qy     422 CAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATCT 481
Db    3321 CAGGAAAGAGATATATATATAAAGTCCTACCACAGGGATGGAAAGGATCGCCAGCAATTT 3380

Qy     482 TCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACCT 541
Db    3381 TTCAACACACAATGAGGCAGGTCTTAGAGCCATTCAGAAAAGCAAACCCAGACGTCATTC 3440

Qy     542 TAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGACA 601
Db    3441 TCGTCCAATATATGGATGATATCTTAATAGCTAGCGACAGGACAGACTTAGAGCATGACA 3500

Qy     602 GGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAGA 661
Db    3501 GAACGGTCCTGCAGTTAAAAGAACTTTTAAATGGCCTAGGATTCTCCACCCCAGATGAGA 3560

Qy     662 AATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGGA 721
Db    3561 AGTTCCAAAAAGACCCCCCATACAAATGGATGGGCTATGAACTATGGCCAACCAAATGGA 3620

Qy     722 AGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAGT 781
Db    3621 AGCTGCAAAAAATACAATTGCCCCAAAAAGAAGTATGGACAGTCAATGACATCCAAAAGC 3680

Qy     782 TAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTCT 841
Db    3681 TAGTAGGTGTCCTAAATTGGGCAGCACAAATCTACCCAGGGATAAAGACCAAACACTTAT 3740

Qy     842 GTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCAG 901
Db    3741 GTAGGCTAATTAGAGGAAAAATGACACTCACGGAAGAAGTACAGTGGACAGAACTAGCAG 3800

Qy     902 AAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTACC 961
Db    3801 AGGCAGAACTAGAAGAGAACAAAATTATCTTGAGCCAGGAACAGGAGGGATGCTATTACC 3860

Qy     962 AAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTATA 1021
Db    3861 AAGAAGAAAAGGAATTAGAAGCAACAGTCCAAAAGGATCAAGACAATCAGTGGACATATA 3920

Qy    1022 AAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACAC 1081
Db    3921 AAATACACCAAGGAGAGAAAATCCTAAAAGTAGGAAAATATGCAAAGATAAAAAATACCC 3980

Qy    1082 ATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATAG 1141
Db    3981 ATACCAATGGGGTCAGATTGTTAGCACATGTAGTTCAAAAAATAGGAAAAGAAGCACTAG 4040

Qy    1142 TGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAGT 1201
Db    4041 TCATTTGGGGACGAATACCAAAATTTCACCTACCAGTAGAAAGAGAAACCTGGGAGCAGT 4100

Qy    1202 GGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCAC 1261
Db    4101 GGTGGGATAACTATTGGCAAGTGACATGGATCCCAGACTGGGACTTCGTATCTACTCCAC 4160

Qy    1262 CGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTATT 1321
Db    4161 CACTGGTCAGGTTAGCATTTAACCTAGTAAAAGATCCCATACCAGGTGAAGAGACCTTCT 4220

Qy    1322 ATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGATA 1381
```

```
Db   4221 ACACAGATGGATCCTGTAATAGGCAATCAAAAGAGGGAAAAGCAGGATATATAACAGATA 4280

Qy   1382 GGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAAG 1441

Db   4281 GAGGGAGAGACAAGGTAAGGATATTGGAGCAAACTACCAATCAGCAAGCAGAATTAGAAG 4340

Qy   1442 CATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAAT 1501

Db   4341 CCTTCGCAATGGCATTAACAGACTCAGGTCCAAAAGCCAATATTATAGTAGACTCACAGT 4400

Qy   1502 ATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAAA 1561

Db   4401 ATGTAATGGGAATAGTAGCGGGCCAGCCAACAGAATCAGAGAGTAAACTAGTAAACCAAA 4460

Qy   1562 TAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAAG 1621

Db   4461 TCATAGAAGAAATGATAAAAAAGGAAACACTCTATGTTGCATGGGTCCCAGCCCACAAAG 4520

Qy   1622 GTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTCT 1681

Db   4521 GCATAGGAGGAAATCAGGAAGTAGATCATTTAGTAAGTCAGGGCATTAGACAAGTATTAT 4580

Qy   1682 TCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAAG 1741

Db   4581 TCCTAGAAAAAATAGAGCCCGCTCAGGAAGAACATGAGAAATATCATAGCAATGTAAAAG 4640

Qy   1742 AATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGTG 1801

Db   4641 AATTATCCCATAAATTTGGACTGCCCAAACTAGTGGCAAGACAAATAGTAAACACATGTG 4700

Qy   1802 ATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACTT 1861

Db   4701 CCCAATGTCAACAGAAAGGGGAAGCTATACATGGGCAAGTAGATGCAGAACTGGGCACTT 4760

Qy   1862 GGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCTA 1921

Db   4761 GGCAAATGGACTGCACACACTTAGAGGGAAAAATCATTATAGTAGCAGTACATGTTGCAA 4820

Qy   1922 GTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTTC 1981

Db   4821 GCGGGTTTATAGAAGCAGAAGTTATCCCACAGGAAACGGGAAGGCAAACAGCACTCTTCC 4880

Qy   1982 TGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAACT 2041

Db   4881 TATTAAAACTGGCCAGTAGGTGGCCAATAACACACCTGCACACAGATAATGGTGCCAACT 4940

Qy   2042 TTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGGG 2101

Db   4941 TCACCTCACAGGAAGTAAAGATGGTAGCGTGGTGGACAGGTATAGAACAATCCTTTGGAG 5000

Qy   2102 TACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAATC 2161

Db   5001 TACCTTACAATCCACAAAGCCAAGGAGTAGTAGAAGCAATGAATCACCACTTAAAAAACC 5060

Qy   2162 AAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTTC 2221

Db   5061 AGATAAGCAGAATTAGAGAGCAGGCAAATACAATGGAAACAATAGTATTAATGGCAGTTC 5120

Qy   2222 ATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTAA 2281

Db   5121 ATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACCCCAGCAGAAAGACTAA 5180

Qy   2282 TTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTTA 2341

Db   5181 TCAATATGATCACCACAGAACAAGAAATACAATTCCTCCACGCAAAAAATTCAAAAATAA 5240

Qy   2342 AAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAGC 2401

Db   5241 AAAATTTCCGGGTCTATTTCAGAGAAGGCAGAGATCAGCTGTGGAAAGGACCTGGGGAAC 5300

Qy   2402 TATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTAC 2461

Db   5301 TACTGTGGAAGGGAGATGGAGCAGTCATAGTCAAGGTAGGGACAGACATAAAAATAGTGC 5360

Qy   2462 CCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGTT 2521

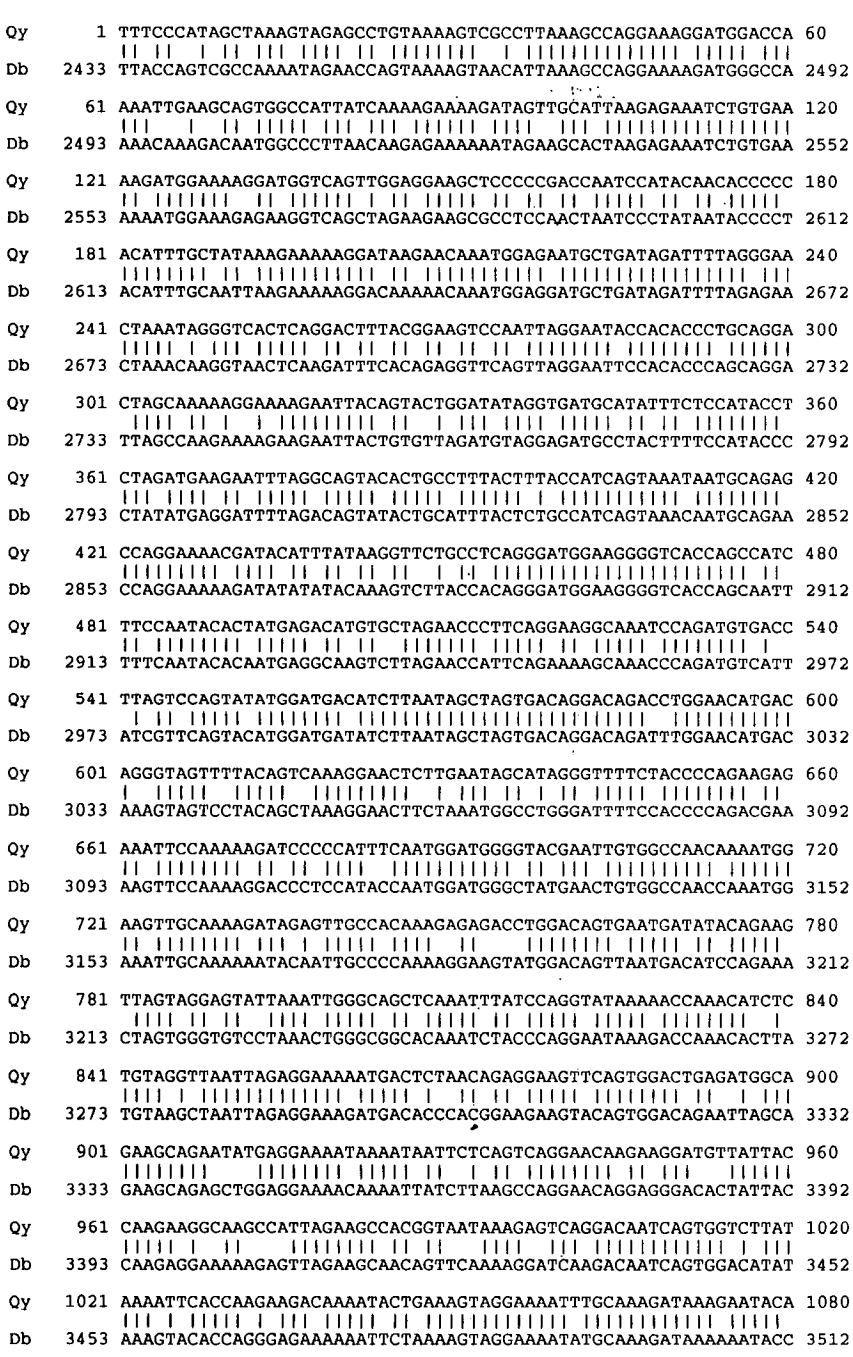
Db   5361 CAAGAAGGAAAGCTAAGATCATCAGAGACTATGGAGGAAGGCGAGAGGTGGATAGTAGTT 5420

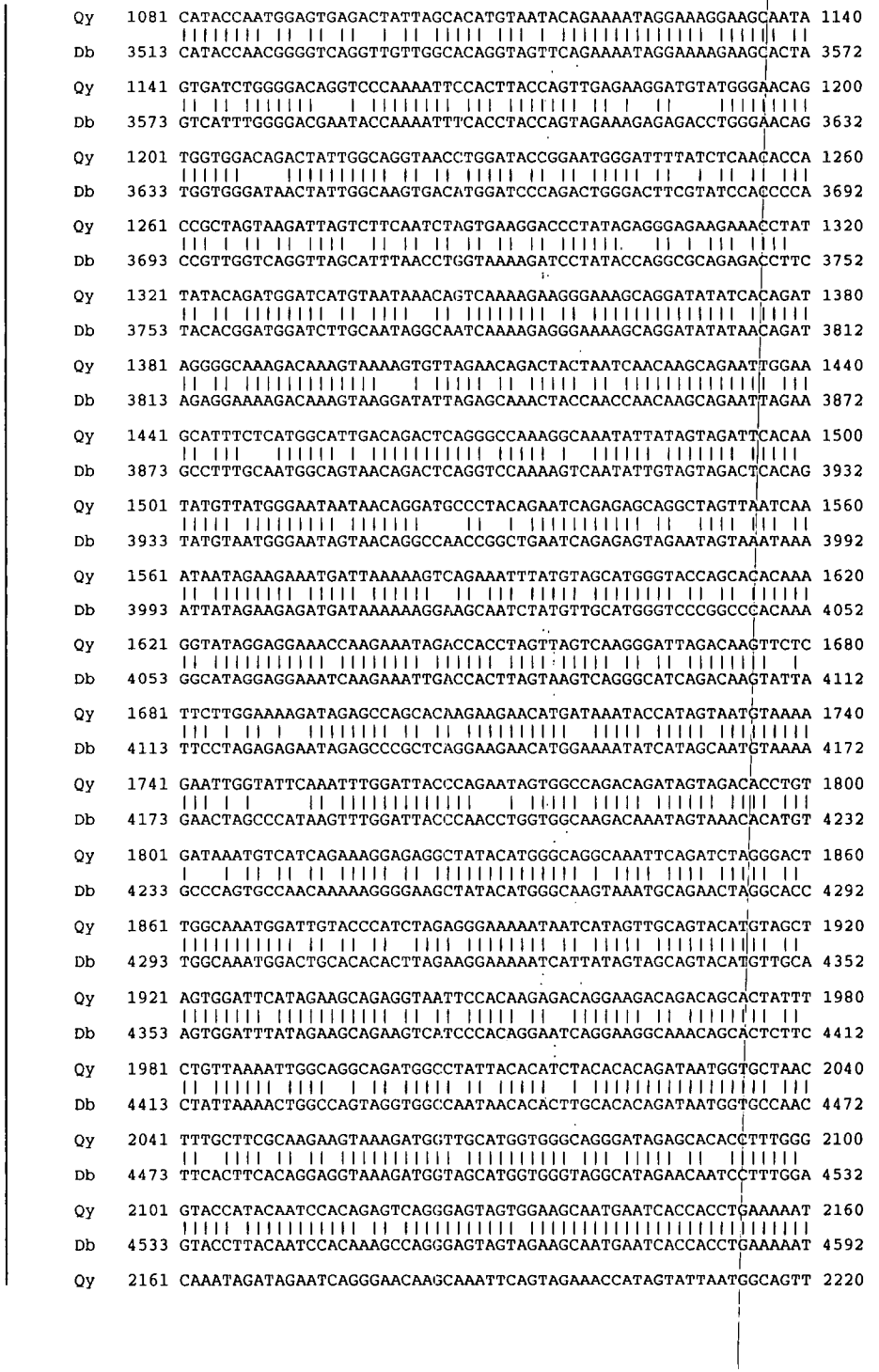
Qy   2522 CCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 2564

Db   5421 CCCACTTGGAGGGTACCAGGGAGGATGGAGAAGTGGCATAGCC 5463
```

RESULT 8
AAN92119
ID AAN92119 standard; DNA; 9693 BP.
XX
AC AAN92119;
XX
DT 17-DEC-2001 (updated)
DT 06-APR-1990 (first entry)
XX
DE Sequence of clone HIV-2 SBL/ISY.
XX
KW HIV-2; proviral clone HIV-2 SBL/ISY;
XX
OS Human immunodeficiency virus 2.
XX
FH Key   Location/Qualifiers
FT CDS   1..9693
FT   /*tag= a
FT CDS   2..9691
FT   /*tag= b
FT CDS   3..9692
FT   /*tag= c
XX
PN USN7331212-N.
XX
PD 29-AUG-1989.
XX
PF 31-MAR-1989; 89US-0331212.
XX
PR 31-MAR-1989; 89US-0331212.
XX
PA (USSH) US DEPT. HEALTH AND HUMAN SERVICES.
XX
PI Franchini G, Wong-Staal F, Gallo R;
XX
DR WPI; 1989-339698/46.
DR P-PSDB; AAP93283. AAP93284. AAP93285.
XX
PT Complete human immunodeficiency type 2 proviral clone - used to generate
PT animal model for function studies of HIV genes in vivo.
XX
PS Disclosure; Fig. 5; 43pp; English.
XX
CC HIV-2 SBL/ISY is a proviral clone of HIV-2. It contains a complete
CC provirus including the cellular flanking sequences. The clone is
CC infectious and cytopathic for some human T cell lines, induces syncytia
CC and infects cell lines in vitro. The clone can be used to generate an
CC animal model for functional studies of HIV genes in vivo, and to develop
CC experimental approaches to prevent retroviral infection in man. It is
CC constructed using DNA from neoplastic human cell line HUT78 freshly
CC infected with the HIV-2 SBL6669 viral isolate. It can be translated in
CC all 3 reading frames.
CC (Note: Revised entry submitted to correct the patent number format of
CC US Government-owned NTIS applications to prevent clashes with ongoing US
CC granted patent numbers. For further information please visit the Derwent
CC web site at www.derwent.com/dwpi/updates/ntis_us.html.)
XX
SQ Sequence 9693 BP; 3315 A; 1947 C; 2412 G; 2019 T; 0 other;

Query Match 67.8%; Score 1738.4; DB 10; Length 9693;
Best Local Similarity 79.9%; Pred. No. 0;
Matches 2048; Conservative 0; Mismatches 516; Indels 0; Gaps 0;





```
Db    4593 CAGATAGAAAGAATTAGAGAGCAGGCAAATACAATGGAAACAATAGTACTAATGGCAGTT 4652

Qy    2221 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTA 2280

Db    4653 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACCCCAGTAGAAAGACTA 4712

Qy    2281 ATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTT 2340

Db    4713 GTCAATATGATCACCACAGAACAAGAAATACAATTCCTCCAAGCAAAAAATTCAAAATTA 4772

Qy    2341 AAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAG 2400

Db    4773 AAAAATTTTCGGGTCTATTTCAGAGAAGGCAGAAATCAACTGTGGCAAGGACCTGGGGAG 4832

Qy    2401 CTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTA 2460

Db    4833 CTACTGTGGAAAGGGGACGGAGCAGTCATAGTCAAGGTAGGGACAGATATAAAAGTAATA 4892

Qy    2461 CCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGT 2520

Db    4893 CCAAGAAGAAAGGCCAAGATCATCAGAGACTATGGACCAAGGCAAGAGATGGATAGCGGT 4952

Qy    2521 TCCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 2564

Db    4953 TCCCACCTGGAGGGTGCCAGGGAGGATGGAGAAATGGCATAGCC 4996
```

RESULT 9
AAC81938
ID AAC81938 standard; DNA; 9663 BP.
XX
AC AAC81938;
XX
DT 28-FEB-2001 (first entry)
XX
DE Packaging vector pCM-ROD(SD36/EM) DNA fragment.
XX
KW Encapsidation; transfer vector; nephrotropic; antiparkinsonian; anti-HIV;
KW cytostatic; gene therapy; transgenic; retroviral packaging;
KW gene delivery; Parkinson's disease; infectious diseases; cancer; ds.
XX
OS Synthetic.
XX
PN WO200040741-A2.
XX
PD 13-JUL-2000.
XX
PF 06-JAN-2000; 2000WO-US00390.
XX
PR 07-JAN-1999; 99US-0115247.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Arya SK;
XX
DR WPI; 2000-475836/41.
XX
PT New lentivirus transfer vector, functionally deleted for a splice donor
PT site and comprising a packaging signal and transgene operably linked to
PT a promoter, for improving encapsidation or transgene RNA and for gene
PT therapy -
XX
PS Example 8; Page 125-130; 143pp; English.
XX
CC This invention describes a novel transfer vector derived from a
CC lentivirus, functionally deleted for a splice donor site (SD), and
CC comprising a packaging signal and transgene operably linked to a
CC promoter. The products of the invention have nephrotropic,
CC antiparkinsonian, anti-HIV, and cytostatic activity and can be used for
CC gene therapy. Encapsidation of transgene RNA is improved using the new
CC retroviral packaging and transfer vectors. The new transfer and packaging
CC vectors are used as gene delivery agents and allows transfer of a
CC transgene into the genome of non-dividing cells. They can be used to
CC create a high-efficiency packaging cell line that provides greatly
CC enhanced packaging of foreign DNA. Individuals suffering from a
CC deficiency in alpha-galactosidase expression, such as Fabry disease can
CC be treated by delivering the vectors to cells in vitro or in vivo.
CC Parkinson's disease, infectious diseases, such as acquired
CC immunodeficiency syndrome and cancers can be treated with the vectors.
CC The non-infective packaging vectors can be used to detect wild-type HIV
CC in biological samples using southern or northern blot assays. The
CC packaging of the vector RNA is maximised, without an increase in the
CC packaging of the viral RNA. Deletion of sequences upstream and
CC downstream of the 5' SD region of the HIV-2 packaging vector results in
CC suppressed encapsidation of the packaging vector genomes without critical
CC loss of gene expression. Functional deletion of the SD site of the
CC transfer vector results in enhanced encapsidation of the transfer
CC vector's genome. HIV-2 packaging vector specifically and faithfully
CC packages its own optimally constructed transfer vector and gives better
CC quality and titre of vector than HIV-1.
XX
SQ Sequence 9663 BP; 3276 A; 1952 C; 2310 G; 2074 T; 51 other;

Query Match 67.6%; Score 1734.2; DB 21; Length 9663;
Best Local Similarity 79.8%; Pred. No. 0;
Matches 2045; Conservative 0; Mismatches 518; Indels 0; Gaps 0;

```
Qy       2 TTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCAA 61

Db    2983 TACCAGTCGCCAAAGTAGAGCCAATAAAAATAATGCTAAAGCCAGGGAAAGATGGACCAA 3042

Qy      62 AATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAAA 121

Db    3043 AACTGAGACAATGGCCCTTAACAAAAGAAAAAATAGAAGCACTAAAAGAAATCTGTGAAA 3102

Qy     122 AGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCCA 181

Db    3103 AAATGGAAAAAGAAGGCCAGCTAGAGGAAGCACCTCCAACTAATCCTTATAATACCCCCA 3162

Qy     182 CATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAAC 241

Db    3163 CATTTGCAATCAAGAAAAAGGACAAAAACAAATGGAGGATGCTAATAGATTTCAGAGAAC 3222

Qy     242 TAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGAC 301

Db    3223 TAAACAAGGTAACTCAAGATTTCACAGAAATTCAGTTAGGAATTCCACACCCAGCAGGGT 3282

Qy     302 TAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCTC 361

Db    3283 TGGCCAAGAAGAGAAGAATTACTGTACTAGATGTAGGGGATGCTTACTTTTCCATACCAC 3342

Qy     362 TAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAGC 421

Db    3343 TACATGAGGACTTTAGACCATATACTGCATTTACTCTACCATCAGTGAACAATGCAGAAC 3402

Qy     422 CAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATCT 481

Db    3403 CAGGAAAAAGATACATATATAAAGTCTTGCCACAGGGATGGAAGGGATCACCAGCAATTT 3462

Qy     482 TCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACCT 541

Db    3463 TTCAACACACAATGAGACAGGTATTAGAACCATTCAGAAAAGCAAACAAGGATGTCATTA 3522

Qy     542 TAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGACA 601

Db    3523 TCATTCAGTACATGGATGATATCTTAATAGCTAGTGACAGGACAGATTTAGAACATGATA 3582

Qy     602 GGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAGA 661

Db    3583 GGGTAGTCCTGCAGCTCAAGGAACTTCTAAATGGCCTAGGATTTTCTACCCCAGATGAGA 3642

Qy     662 AATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGGA 721

Db    3643 AGTTCCAAAAAGACCCTCCATACCACTGGATGGGCTATGAACTATGGCCAACTAAATGGA 3702
```

```
Qy    722 AGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAGT 781
Db   3703 AGTTGCAGAAAATACAGTTGCCCCAAAAAGAAATATGGACAGTCAATGACATCCAGAAGC 3762

Qy    782 TAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTCT 841
Db   3763 TAGTGGGTGTCCTAAATTGGGCAGCACAACTCTACCCAGGGATAAAGACCAAACACTTAT 3822

Qy    842 GTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCAG 901
Db   3823 GTAGGTTAATCAGAGGAAAAATGACACTCACAGAAGAAGTACAGTGGACAGAATTAGCAG 3882

Qy    902 AAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTACC 961
Db   3883 AAGCAGAGCTAGAAGAAAACAGAATTATCCTAAGCCAGGAACAAGAGGGACACTATTACC 3942

Qy    962 AAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTATA 1021
Db   3943 AAGAAGAAAAAGAGCTAGAAGCAACAGTCCAAAAGGATCAAGAGAATCAGTGGACATATA 4002

Qy   1022 AAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACAC 1081
Db   4003 AAATACACCAGGAAGAAAAAATTCTAAAAGTAGGAAAATATGCAAAGGTGAAAAACACCC 4062

Qy   1082 ATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATAG 1141
Db   4063 ATACCAATGGAATCAGATTGTTAGCACAGGTAGTTCAGAAAATAGGAAAAGAAGCACTAG 4122

Qy   1142 TGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAGT 1201
Db   4123 TCATTTGGGGACGAATACCAAAATTTCACCTACCAGTAGAGAGAGAAATCTGGGAGCAGT 4182

Qy   1202 GGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCAC 1261
Db   4183 GGTGGGATAACTACTGGCAAGTGACATGGATCCCAGACTGGGACTTCGTGTCTACCCCAC 4242

Qy   1262 CGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTATT 1321
Db   4243 CACTGGTCAGGTTAGCGTTTAACCTGGTAGGGGATCCTATACCAGGTGCAGAGACCTTCT 4302

Qy   1322 ATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGATA 1381
Db   4303 ACACAGATGGATCCTGCAATAGGCAATCAAAAGAAGGAAAAGCAGGATATGTAACAGATA 4362

Qy   1382 GGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAAG 1441
Db   4363 GAGGGAAAGACAAGGTAAAGAAACTAGAGCAAACTACCAATCAGCAAGCAGAACTAGAAG 4422

Qy   1442 CATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAAT 1501
Db   4423 CCTTTGCGATGGCACTAACAGACTCGGGTCCAAAAGTTAATATTATAGTAGACTCACAGT 4482

Qy   1502 ATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAAA 1561
Db   4483 ATGTAATGGGGATCAGTGCAAGCCAACCAACAGAGTCAGAAAGTAAAATAGTGAACCAGA 4542

Qy   1562 TAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAAG 1621
Db   4543 TCATAGAAGAAATGATAAAAAAGGAAGCAATCTATGTTGCATGGGTCCCAGCCCACAAAG 4602

Qy   1622 GTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTCT 1681
Db   4603 GCATAGGGGGAAACCAGGAAGTAGATCATTTAGTGAGTCAGGGTATCAGACAAGTGTTGT 4662

Qy   1682 TCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAAG 1741
Db   4663 TCCTGGAAAAAATAGAGCCCGCTCAGGAAGAACATGAAAAATATCATAGCAATGTAAAAG 4722

Qy   1742 AATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGTG 1801
Db   4723 AACTGTCTCATAAATTTGGAATACCCAATTTAGTGGCAAGGCAAATAGTAAACTCATGTG 4782

Qy   1802 ATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACTT 1861

Db   4783 CCCAATGTCAACAGAAAGGGGAAGCTATACATGGGCAAGTAAATGCAGAACTAGGCACTT 4842

Qy   1862 GGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCTA 1921
Db   4843 GGCAAATGGACTGCACACATTTAGAAGGAAAGATCATTATAGTAGCAGTACATGTTGCAA 4902

Qy   1922 GTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTTC 1981
Db   4903 GTGGATTTATAGAAGCAGAAGTCATCCCACAGGAATCAGGAAGACAAACAGCACTCTTCC 4962

Qy   1982 TGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAACT 2041
Db   4963 TATTGAAACTGGCAAGTAGGTGGCCAATAACACACTTGCATACAGATAATGGTGCCAACT 5022

Qy   2042 TTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGGG 2101
Db   5023 TCACTTCACAGGAGGTGAAGATGGTAGCATGGTGGATAGGTATAGAACAATCCTTTGGAG 5082

Qy   2102 TACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAATC 2161
Db   5083 TACCTTACAATCCACAGAGCCAAGGAGTAGTAGAAGCAATGAATCACCATCTAAAAAACC 5142

Qy   2162 AAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTTC 2221
Db   5143 AAATAAGTAGAATCAGAGAACAGGCAAATACAATAGAAACAATAGTACTAATGGCAATTC 5202

Qy   2222 ATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTAA 2281
Db   5203 ATTGCATGAATTTTAAAAGAAGGGGGGGAATAGGGGATATGACTCCATCAGAAAGATTAA 5262

Qy   2282 TTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTTA 2341
Db   5263 TCAATATGATCACCACAGAACAAGAGATACAATTCCTCCAAGCCAAAAATTCAAAATTAA 5322

Qy   2342 AAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAGC 2401
Db   5323 AAGATTTTCGGGTCTATTTCAGAGAAGGCAGAGATCAGTTGTGGAAAGGACCTGGGGAAC 5382

Qy   2402 TATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTAC 2461
Db   5383 TACTGTGGAAAGGAGAAGGAGCAGTCCTAGTCAAGGTAGGAACAGACATAAAAATAATAC 5442

Qy   2462 CCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGTT 2521
Db   5443 CAAGAAGGAAAGCCAAGATCATCAGAGACTATGGAGGAAGACAAGAGATGGATAGTGGTT 5502

Qy   2522 CCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 2564
Db   5503 CCCACCTGGAGGGTGCCAGGGAGGATGGAGAAATGGCATAGCC 5545
```

RESULT 10
AAC81937
ID AAC81937 standard; DNA; 9726 BP.
XX
AC AAC81937;
XX
DT 28-FEB-2001 (first entry)
XX
DE Packaging vector pROD(SD36/EM) DNA fragment.
XX
KW Encapsidation; transfer vector; nephrotropic; antiparkinsonian; anti-HIV;
KW cytostatic; gene therapy; transgenic; retroviral packaging;
KW gene delivery; Parkinson's disease; infectious diseases; cancer; ds.
XX
OS Synthetic.
XX
PN WO200040741-A2.
XX
PD 13-JUL-2000.
XX
PF 06-JAN-2000; 2000WO-US00390.

XX
PR 07-JAN-1999; 99US-0115247.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Arya SK;
XX
DR WPI; 2000-475836/41.
XX
PT New lentivirus transfer vector, functionally deleted for a splice donor
PT site and comprising a packaging signal and transgene operably linked to
PT a promoter, for improving encapsidation or transgene RNA and for gene
PT therapy -
XX
PS Example 6; Page 119-124; 143pp; English.
XX
CC This invention describes a novel transfer vector derived from a
CC lentivirus, functionally deleted for a splice donor site (SD), and
CC comprising a packaging signal and transgene operably linked to a
CC promoter. The products of the invention have nephrotropic,
CC antiparkinsonian, anti-HIV, and cytostatic activity and can be used for
CC gene therapy. Encapsidation of transgene RNA is improved using the new
CC retroviral packaging and transfer vectors. The new transfer and packaging
CC vectors are used as gene delivery agents and allows transfer of a
CC transgene into the genome of non-dividing cells. They can be used to
CC create a high-efficiency packaging cell line that provides greatly
CC enhanced packaging of foreign DNA. Individuals suffering from a
CC deficiency in alpha-galactosidase expression, such as Fabry disease can
CC be treated by delivering the vectors to cells in vitro or in vivo.
CC Parkinson's disease, infectious diseases, such as acquired
CC immunodeficiency syndrome and cancers can be treated with the vectors.
CC The non-infective packaging vectors can be used to detect wild-type HIV
CC in biological samples using southern or northern blot assays. The
CC packaging of the vector RNA is maximised, without an increase in the
CC packaging of the viral RNA. Deletion of sequences upstream and
CC downstream of the 5' SD region of the HIV-2 packaging vector results in
CC suppressed encapsidation of the packaging vector genomes without critical
CC loss of gene expression. Functional deletion of the SD site of the
CC transfer vector results in enhanced encapsidation of the transfer
CC vector's genome. HIV-2 packaging vector specifically and faithfully
CC packages its own optimally constructed transfer vector and gives better
CC quality and titre of vector than HIV-1.
XX
SQ Sequence 9726 BP; 3265 A; 1937 C; 2354 G; 1968 T; 202 other;

Query Match 67.6%; Score 1734.2; DB 21; Length 9726;
Best Local Similarity 79.8%; Pred. No. 0;
Matches 2045; Conservative 0; Mismatches 518; Indels 0; Gaps 0;

```
Qy       2 TTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCAA 61
Db    2379 TACCAGTCGCCAAAGTAGAGCCAATAAAAATAATGCTAAAGCCAGGGAAAGATGGACCAA 2438

Qy      62 AATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAAA 121
Db    2439 AACTGAGACAATGGCCCTTAACAAAAGAAAAAATAGAAGCACTAAAAGAAATCTGTGAAA 2498

Qy     122 AGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCCA 181
Db    2499 AAATGGAAAAAGAAGGCCAGCTAGAGGAAGCACCTCCAACTAATCCTTATAATACCCCCA 2558

Qy     182 CATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAAC 241
Db    2559 CATTTGCAATCAAGAAAAAGGACAAAAACAAATGGAGGATGCTAATAGATTTCAGAGAAC 2618

Qy     242 TAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGAC 301
Db    2619 TAAACAAGGTAACTCAAGATTTCACAGAAATTCAGTTAGGAATTCCACACCCAGCAGGGT 2678

Qy     302 TAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCTC 361
Db    2679 TGGCCAAGAAGAGAAGAATTACTGTACTAGATGTAGGGGATGCTTACTTTTCCATACCAC 2738

Qy     362 TAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAGC 421
Db    2739 TACATGAGGACTTTAGACCATATACTGCATTTACTCTACCATCAGTGAACAATGCAGAAC 2798

Qy     422 CAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATCT 481
Db    2799 CAGGAAAAAGATACATATATAAAGTCTTGCCACAGGGATGGAAGGGATCACCAGCAATTT 2858

Qy     482 TCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACCT 541
Db    2859 TTCAACACACAATGAGACAGGTATTAGAACCATTCAGAAAAGCAAACAAGGATGTCATTA 2918

Qy     542 TAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGACA 601
Db    2919 TCATTCAGTACATGGATGATATCTTAATAGCTAGTGACAGGACAGATTTAGAACATGATA 2978

Qy     602 GGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAGA 661
Db    2979 GGGTAGTCCTGCAGCTCAAGGAACTTCTAAATGGCCTAGGATTTTCTACCCCAGATGAGA 3038

Qy     662 AATTCCAAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGGA 721
Db    3039 AGTTCCAAAAAGACCCTCCATACCACTGGATGGGCTATGAACTATGGCCAACTAAATGGA 3098

Qy     722 AGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAGT 781
Db    3099 AGTTGCAGAAAATACAGTTGCCCCAAAAAGAAATATGGACAGTCAATGACATCCAGAAGC 3158

Qy     782 TAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTCT 841
Db    3159 TAGTGGGTGTCCTAAATTGGGCAGCACAACTCTACCCAGGGATAAAGACCAAACACTTAT 3218

Qy     842 GTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCAG 901
Db    3219 GTAGGTTAATCAGAGGAAAAATGACACTCACAGAAGAAGTACAGTGGACAGAATTAGCAG 3278

Qy     902 AAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTACC 961
Db    3279 AAGCAGAGCTAGAAGAAAACAGAATTATCCTAAGCCAGGAACAAGAGGGACACTATTACC 3338

Qy     962 AAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTATA 1021
Db    3339 AAGAAGAAAAAGAGCTAGAAGCAACAGTCCAAAAGGATCAAGAGAATCAGTGGACATATA 3398

Qy    1022 AAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACAC 1081
Db    3399 AAATACACCAGGAAGAAAAAATTCTAAAAGTAGGAAAATATGCAAAGGTGAAAAACACCC 3458

Qy    1082 ATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATAG 1141
Db    3459 ATACCAATGGAATCAGATTGTTAGCACAGGTAGTTCAGAAAATAGGAAAAGAAGCACTAG 3518

Qy    1142 TGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAGT 1201
Db    3519 TCATTTGGGGACGAATACCAAAATTTCACCTACCAGTAGAGAGAGAAATCTGGGAGCAGT 3578

Qy    1202 GGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCAC 1261
Db    3579 GGTGGGATAACTACTGGCAAGTGACATGGATCCCAGACTGGGACTTCGTGTCTACCCCAC 3638

Qy    1262 CGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTATT 1321
Db    3639 CACTGGTCAGGTTAGCGTTTAACCTGGTAGGGGATCCTATACCAGGTGCAGAGACCTTCT 3698

Qy    1322 ATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGATA 1381
Db    3699 ACACAGATGGATCCTGCAATAGGCAATCAAAAGAAGGAAAAGCAGGATATGTAACAGATA 3758

Qy    1382 GGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAAG 1441
Db    3759 GAGGGAAAGACAAGGTAAAGAAACTAGAGCAAACTACCAATCAGCAAGCAGAACTAGAAG 3818

Qy    1442 CATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAAT 1501
```

```
Db   3819 CCTTTGCGATGGCACTAACAGACTCGGGTCCAAAAGTTAATATTATAGTAGACTCACAGT 3878

Qy   1502 ATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAAA 1561

Db   3879 ATGTAATGGGGATCAGTGCAAGCCAACCAACAGAGTCAGAAAGTAAAATAGTGAACCAGA 3938

Qy   1562 TAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAAG 1621

Db   3939 TCATAGAAGAAATGATAAAAAAGGAAGCAATCTATGTTGCATGGGTCCCAGCCCACAAAG 3998

Qy   1622 GTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTCT 1681

Db   3999 GCATAGGGGGAAACCAGGAAGTAGATCATTTAGTGAGTCAGGGTATCAGACAAGTGTTGT 4058

Qy   1682 TCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAAG 1741

Db   4059 TCCTGGAAAAAATAGAGCCCGCTCAGGAAGAACATGAAAAATATCATAGCAATGTAAAAG 4118

Qy   1742 AATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGTG 1801

Db   4119 AACTGTCTCATAAATTTGGAATACCCAATTTAGTGGCAAGGCAAATAGTAAACTCATGTG 4178

Qy   1802 ATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACTT 1861

Db   4179 CCCAATGTCAACAGAAAGGGGAAGCTATACATGGGCAAGTAAATGCAGAACTAGGCACTT 4238

Qy   1862 GGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCTA 1921

Db   4239 GGCAAATGGACTGCACACATTTAGAAGGAAAGATCATTATAGTAGCAGTACATGTTGCAA 4298

Qy   1922 GTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTTC 1981

Db   4299 GTGGATTTATAGAAGCAGAAGTCATCCCACAGGAATCAGGAAGACAAACAGCACTCTTCC 4358

Qy   1982 TGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAACT 2041

Db   4359 TATTGAAACTGGCAAGTAGGTGGCCAATAACACACTTGCATACAGATAATGGTGCCAACT 4418

Qy   2042 TTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGGG 2101

Db   4419 TCACTTCACAGGAGGTGAAGATGGTAGCATGGTGGATAGGTATAGAACAATCCTTTGGAG 4478

Qy   2102 TACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAATC 2161

Db   4479 TACCTTACAATCCACAGAGCCAAGGAGTAGTAGAAGCAATGAATCACCATCTAAAAAACC 4538

Qy   2162 AAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTTC 2221

Db   4539 AAATAAGTAGAATCAGAGAACAGGCAAATACAATAGAAACAATAGTACTAATGGCAATTC 4598

Qy   2222 ATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTAA 2281

Db   4599 ATTGCATGAATTTTAAAAGAAGGGGGGGAATAGGGGATATGACTCCATCAGAAAGATTAA 4658

Qy   2282 TTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTTA 2341

Db   4659 TCAATATGATCACCACAGAACAAGAGATACAATTCCTCCAAGCCAAAAATTCAAAATTAA 4718

Qy   2342 AAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAGC 2401

Db   4719 AAGATTTTCGGGTCTATTTCAGAGAAGGCAGAGATCAGTTGTGGAAAGGACCTGGGGAAC 4778

Qy   2402 TATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTAC 2461

Db   4779 TACTGTGGAAAGGAGAAGGAGCAGTCCTAGTCAAGGTAGGAACAGACATAAAAATAATAC 4838

Qy   2462 CCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGTT 2521

Db   4839 CAAGAAGGAAAGCCAAGATCATCAGAGACTATGGAGGAAGACAAGAGATGGATAGTGGTT 4898

Qy   2522 CCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 2564

Db   4899 CCCACCTGGAGGGTGCCAGGGAGGATGGAGAAATGGCATAGCC 4941
```

```
RESULT 11
AAQ20616
ID   AAQ20616 standard; DNA; 9672 BP.
XX
AC   AAQ20616;
XX
DT   19-MAY-1992  (first entry)
XX
DE   ROD HIV-2 isolate complete genome.
XX
KW   Human immunodeficiency virus; AIDS; ss.
XX
OS   Human immunodeficiency virus-2 ROD isolate.
XX
FH   Key             Location/Qualifiers
FT   CDS             546..2111
FT                   /*tag= a
FT                   /note= "GAG gene"
FT   Misc_feature    546..950
FT                   /*tag= b
FT                   /note= "GAG p 16"
FT   Misc_feature    951..1640
FT                   /*tag= c
FT                   /note= "GAG p 26"
FT   Misc_feature    1701..2111
FT                   /*tag= d
FT                   /note= "GAG p 12"
FT   Misc_feature    1829..4936
FT                   /*tag= e
FT                   /note= "polymerase"
FT   Misc_feature    4869..5513
FT                   /*tag= f
FT                   /note= "Q protein"
FT   Misc_feature    5682..5996
FT                   /*tag= g
FT                   /note= "R protein"
FT   Misc_feature    5344..5679
FT                   /*tag= h
FT                   /note= "X protein"
FT   Misc_feature    5682..5996
FT                   /*tag= i
FT                   /note= "Y protein"
FT   Misc_feature    6147..8720
FT                   /*tag= j
FT                   /note= "Env protein"
FT   Misc_feature    8557..9324
FT                   /*tag= k
FT                   /note= "F protein"
FT   Exon            5845..6140
FT                   /*tag= l
FT                   /note= "TAT gene exon 1"
FT   Exon            8307..8400
FT                   /*tag= m
FT                   /note= "TAT gene exon 2"
FT   Exon            6071..6140
FT                   /*tag= n
FT                   /note= "ART gene exon 1"
FT   Exon            8307..8536
FT                   /*tag= o
FT                   /note= "ART gene exon 2"
FT   LTR             1..173
FT                   /*tag= p
FT                   /note= "LTR:R 1"
FT   LTR             9498..9671
FT                   /*tag= q
FT                   /note= "LTR:R 2"
FT   Misc_feature    174..299
FT                   /*tag= r
FT                   /note= "U5"
FT   Misc_feature    8942..9497
```

```
FT                   /*tag=  s
FT                   /note= "U3"
XX
PN   US5079342-A.
XX
PD   07-JAN-1992.
XX
PF   11-FEB-1987;   87US-0U13477.
XX
PR   11-FEB-1987;   87US-0013477.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Alizon M,  Montagnier L,  Geutard D,  Clavel F,  Sonigo P,  Guyader M;
XX
DR   WPI; 1992-041067/05.
DR   P-PSDB; AAR20596, AAR20597, AAR20598, AAR20599, AAR20600, AAR20601,
DR   AAR20602, AAR20603, AAR20604, AAR20605, AAR20606, AAR20607.
XX
PT   Peptide(s) corresp. to HIV-2 amino acid sequences - used in
PT   diagnosis in vaccines and in prodn. of antibodies for diagnosis
XX
PS   Disclosure; Page 12; 30pp; English.
XX
CC   The sequence is that of the complete genomic sequence of the ROD HIV-2
CC   isolate. Its genes and their major expression products are indicated
CC   in the features. It was isolated from a CEM cell line infected with
CC   the ROD isolate.
XX
SQ   Sequence 9672 BP; 3314 A; 1981 C; 2394 G; 1983 T; 0 other;

  Query Match             67.4%;  Score 1727.8;  DB 13;  Length 9672;
  Best Local Similarity   79.6%;  Pred. No. 0;
  Matches 2041;  Conservative    0;  Mismatches  522;  Indels    0;  Gaps    0;

Qy       2 TTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCAA 61
Db    2379 TACCAGTCGCCAAAGTAGAGCCAATAAAAATAATGCTAAAGCCAGGGAAAGATGGACCAA 2438

Qy      62 AATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAAA 121
Db    2439 AACTGAGACAATGGCCCTTAACAAAAGAAAAAATAGAAGCACTAAAAGAAATCTGTGAAA 2498

Qy     122 AGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCCA 181
Db    2499 AAATGGAAAAAGAAGGCCAGCTAGAGGAAGCACCTCCAACTAATCCTTATAATACCCCCA 2558

Qy     182 CATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAAC 241
Db    2559 CATTTGCAATCAAGAAAAAGGACAAAAACAAATGGAGGATGCTAATAGATTTCAGAGAAC 2618

Qy     242 TAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGAC 301
Db    2619 TAAACAAGGTAACTCAAGATTTCACAGAAATTCAGTTAGGAATTCCACACCCAGCAGGGT 2678

Qy     302 TAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCTC 361
Db    2679 TGGCCAAGAAGAGAAGAATTACTGTACTAGATGTAGGGGATGCTTACTTTTCCATACCAC 2738

Qy     362 TAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAGC 421
Db    2739 TACATGAGGACTTTAGACCATATACTGCATTTACTCTACCATCAGTGAACAATGCAGAAC 2798

Qy     422 CAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATCT 481
Db    2799 CAGGAAAAAGATACATATATAAAGTCTTGCCACAGGGATGGAAGGGATCACCAGCAATTT 2858

Qy     482 TCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACCT 541
Db    2859 TTCAACACACAATGAGACAGGTATTAGAACCATTCAGAAAAGCAAACAAGGATGTCATTA 2918

Qy     542 TAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGACA 601
Db    2919 TCATTCAGTACATGGATGATATCTTAATAGCTAGTGACAGGACAGATTTAGAACATGATA 2978

Qy     602 GGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAGA 661
Db    2979 GGGTAGTCCTGCAGCTCAAGGAACTTCTAAATGGCCTAGGATTTTCTACCCCAGATGAGA 3038

Qy     662 AATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGGA 721
Db    3039 AGTTCCAAAAAGACCCTCCATACCACTGGATGGGCTATGAACTATGGCCAACTAAATGGA 3098

Qy     722 AGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAGT 781
Db    3099 AGTTGCAGAAAATACAGTTGCCCCAAAAAGAAATATGGACAGTCAATGACATCCAGAAGC 3158

Qy     782 TAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTCT 841
Db    3159 TAGTGGGTGTCCTAAATTGGGCAGCACAACTCTACCCAGGGATAAAGACCAAACACTTAT 3218

Qy     842 GTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCAG 901
Db    3219 GTAGGTTAATCAGAGGAAAAATGACACTCACAGAAGAAGTACAGTGGACAGAATTAGCAG 3278

Qy     902 AAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTACC 961
Db    3279 AAGCAGAGCTAGAAGAAAACAGAATTATCCTAAGCCAGGAACAAGAGGGACACTATTACC 3338

Qy     962 AAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTATA 1021
Db    3339 AAGAAGAAAAAGAGCTAGAAGCAACAGTCCAAAAGGATCAAGAGAATCAGTGGACATATA 3398

Qy    1022 AAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACAC 1081
Db    3399 AAATACACCAGGAAGAAAAAATTCTAAAAGTAGGAAAATATGCAAAGGTGAAAAACACCC 3458

Qy    1082 ATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATAG 1141
Db    3459 ATACCAATGGAATCAGATTGTTAGCACAGGTAGTTCAGAAAATAGGAAAAGAAGCACTAG 3518

Qy    1142 TGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAGT 1201
Db    3519 TCATTTGGGGACGAATACCAAAATTTCACCTACCAGTAGAGAGAGAAATCTGGGAGCAGT 3578

Qy    1202 GGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCAC 1261
Db    3579 GGTGGGATAACTACTGGCAAGTGACATGGATCCCAGACTGGGACTTCGTGTCTACCCCAC 3638

Qy    1262 CGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTATT 1321
Db    3639 CACTGGTCAGGTTACCCTTTAACCTGGTAGGGGATCCTATACCAGGTGCAGAGACCTTCT 3698

Qy    1322 ATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGATA 1381
Db    3699 ACACACATGGATCCTGCAATAGGCAATCAAAAGAAGGAAAAGCAGGATATGTAACAGATA 3758

Qy    1382 GGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAAG 1441
Db    3759 GAGGGAAAGACAAGGTAAAGAAACTAGAGCAAACTACCAATCAGCAAGCAGAACTAGAAG 3818

Qy    1442 CATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAAT 1501
Db    3819 CCTTTGCGATGGCACTAACAGACTCGGGTCCAAAAGTTAATATTATAGTAGACTCACAGT 3878

Qy    1502 ATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAAA 1561
Db    3879 ATGTAATGGGGATCAGTGCAAGCCAACCAACAGAGTCAGAAAGTAAAATAGTGAACCAGA 3938

Qy    1562 TAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAAG 1621
Db    3939 TCATAGAAGAAATGATAAAAAAGGAAGCAATCTATGTTGCATGGGTCCCAGCCCACAAAG 3998

Qy    1622 GTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTCT 1681
Db    3999 GCATAGGGGGAAACCAGGAAGTAGATCATTTAGTGAGTCAGGGTATCAGACAAGTGTTGT 4058
```

```
Qy   1682 TCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAAG 1741
          || ||||||| |||||||| || || |||||||||| ||||| ||||| ||||||||||
Db   4059 TCCTGGAAAAAATAGAGCCCGCTCAGGAAGAACATGAAAAATATCATAGCAATGTAAAAG 4118

Qy   1742 AATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGTG 1801
          || ||      ||||||||| ||||||   ||||||| || || |||||| || | ||||
Db   4119 AACTGTCTCATAAATTTGGAATACCCAATTTAGTGGCAAGGCAAATAGTAAACTCATGTG 4178

Qy   1802 ATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACTT 1861
             ||||||| |||||||| || |||||||||||||| | |||| |||| ||||| ||||
Db   4179 CCCAATGTCAACAGAAAGGGGAAGCTATACATGGGCAAGTAAATGCAGAACTAGGCACTT 4238

Qy   1862 GGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCTA 1921
          |||||||||| || || ||| |||| | ||| || || ||||| ||||||||||| || |
Db   4239 GGCAAATGGACTGCACACATTTAGAAGCAAAGATCATTATAGTAGCAGTACATGTTGCAA 4298

Qy   1922 GTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTTC 1981
          ||||||| |||||||||| || || ||||| ||  |||||||||| |||||||| || |
Db   4299 GTGGATTTATAGAAGCAGAAGTCATCCCACAGGAATCAGGAAGACAAACAGCACTCTTCC 4358

Qy   1982 TGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAACT 2041
          | || ||| ||||| | || ||||| || |||||  | || ||||||||||||| ||||
Db   4359 TATTGAAACTGGCAAGTAGGTGGCCAATAACACACTTGCATACAGATAATGGTGCCAACT 4418

Qy   2042 TTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGGG 2101
          |  |||| || || || |||||||| |||||||||  ||| ||||| ||  |||||||  |
Db   4419 TCACTTCACAGGAGGTGAAGATGGTAGCATGGTGGATAGGTATAGAACAATCCTTTGGAG 4478

Qy   2102 TACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAATC 2161
          |||| |||||||||||||| || |||||||| |||||||||||||||| || ||||| |
Db   4479 TACCTTACAATCCACAGAGCCAAGGAGTAGTAGAAGCAATGAATCACCATCTAAAAAACC 4538

Qy   2162 AAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTTC 2221
          |||||  |||||||| ||||| |||||| || ||||||| |||||| |||||||| |||
Db   4539 AAATAAGTAGAATCAGAGAACAGGCAAATACAATAGAAACAATAGTACTAATGGCAATTC 4598

Qy   2222 ATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTAA 2281
          |||||||||||||||||||||||| |||||||||||||||||||| ||||||||||||
Db   4599 ATTGCATGAATTTTAAAAGAAGGGGGGGAATAGGGGATATGACTCCATCAGAAAGATTAA 4658

Qy   2282 TTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTTA 2341
          | || |||||||| |||||||||||||||||||| |  ||| | ||||| |||||||| |
Db   4659 TCAATATGATCACCACAGAACAAGAGATACAATTCCTCCAAGCCAAAAATTCAAAATTAA 4718

Qy   2342 AAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAGC 2401
          || |||||||||||||| |||||||||| |||||| |||||| ||||| || || |
Db   4719 AAGATTTTCGGGTCTATTTCAGAGAAGGCACAGATCAGTTGTGGAAAGGACCTGGGGAAC 4778

Qy   2402 TATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTAC 2461
          || |||||||||| |||||||||||  |  | ||||||| |||||||| ||  || |||
Db   4779 TACTGTGGAAAGGAGAAGGAGCAGTCCTAGTCAAGGTAGGAACAGACATAAAAATAATAC 4838

Qy   2462 CCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGTT 2521
          | ||||| || || || ||  ||| ||| |||||||||| || ||||| |||||||  |||
Db   4839 CAAGAAGGAAAGCCAAGATCCTCAGAGACTATGGAGGAAGACAAGAGATGGATAGTGGTT 4898

Qy   2522 CCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 2564
          ||||| |||||| | || | |||| | ||||  ||||||||||
Db   4899 CCCACCTGGAGGGTGCCAGGGAGGATGGAGAAATGGCATAGCC 4941


RESULT 12
AAN80890
ID   AAN80890 standard; cDNA; 9633 BP.
XX
AC   AAN80890;
XX
DT   07-NOV-1990  (first entry)
XX
DE   Sequence of cDNA clone HIV-2 SBL/ISY of HIV related retrovirus strain.
XX
KW   HIV vaccine; HIV strain SBL-6669-85; ss.
XX
OS   Human immunodeficiency virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..9633
FT                   /*tag= a
FT                   /note="SQ in AAP81769"
FT   CDS             2..9631
FT                   /*tag= b
FT                   /note="SQ in AAP81770"
FT   CDS             3..9632
FT                   /*tag= c
FT                   /note="SQ in AAP81771"
XX
PN   WO8808449-A.
XX
PD   03-NOV-1988.
XX
PF   28-APR-1988;   88WO-SE00218.
XX
PR   28-APR-1987;   87SE-0001765.
XX
PA   (SBLS-) SBL STATENS BEKTERI.
XX
PI   Albert J,  Biberfeld G,  Fenyo EM,  Norrby E;
XX
DR   WPI; 1988-322769/45.
DR   P-PSDB; AAP81769, AAP81770, AAP81771.
XX
PT   HIV related human retro-virus strain -
PT   used for obtaining antigens for assays and vaccines and for
PT   prodn. of antibodies for assays
XX
PS   Claim 7; Fig 4; 28pp; English.
XX
CC   cDNA clone (HIV-2 SBL/ISY) characterised by having the primary nucleotide
CC   sequence in n80890 or a degenerate or a part thereof, esp. a clone having
CC   capacity to give prodn. of infectious virus particles is claimed.
CC   HIV-2 SBL/ISY represents the complete genome of the virus SBL-6699
CC   (=SBL-6689-85). The proviral DNA was obtd. from a genomic library
CC   constructed from HUT-78 cells infected with SBL-6669-85 virus using
CC   the lambda-phage vector EMBL-3. SBL-6669-85 was isolated from lymphocytes
CC   of a West African woman. Protection is requested for the entire genome
CC   disclosed in n80890 and for parts thereof, and corresp. to various genes
CC   such as the gag gene (corresp. to nucleotides 547 to 2106), the pol gene
CC   (nucleotides 1827-4931) and the env gene (nucleotides 6144 to 8682), the
CC   corresp. AA sequences and parts thereof and various products derived
CC   therefrom, or use thereof, such as clones prepd. by recombinant vector
CC   method, HIV test devices and methods.
XX
SQ   Sequence 9633 BP; 3288 A; 1943 C; 2394 G; 2008 T; 0 other;

  Query Match             67.2%;  Score 1724.2;  DB 9;  Length 9633;
  Best Local Similarity   79.8%;  Pred. No. 0;
  Matches 2032;  Conservative    0;  Mismatches  513;  Indels    0;  Gaps    0;

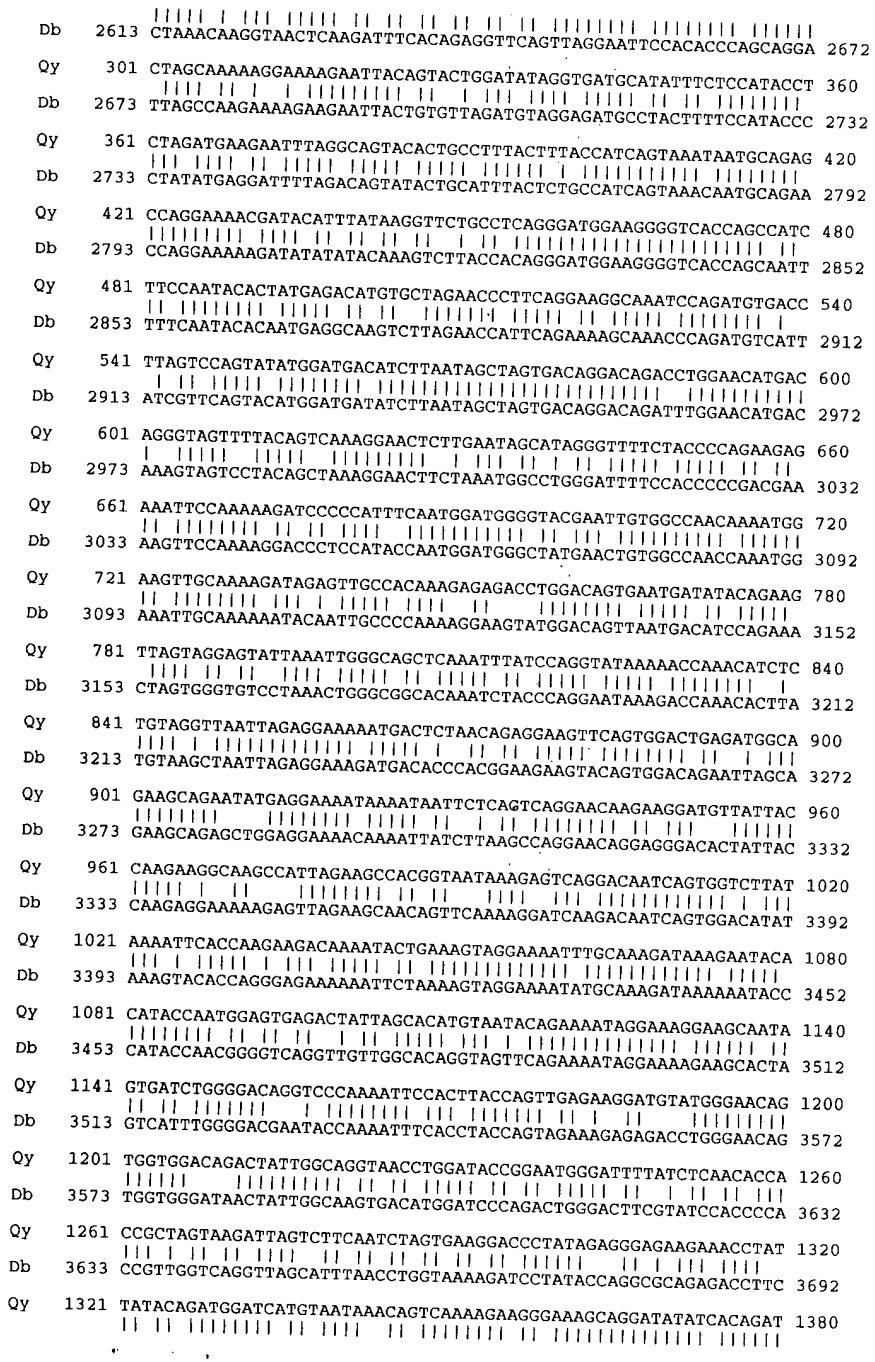
Qy      1 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 60
          || ||  | || ||| |||| || ||||||||  | ||||||||||||||| |||||| |||
Db   2373 TTACCAGTCGCCAAAATAGAACCAGTAAAAGTAACATTAAAGCCAGGAAAAGATGGGCCA 2432

Qy     61 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 120
          |||   |  || ||||| ||| ||| |||||| ||||  ||| ||||||||||||||||||
Db   2433 AAACAAAGACAATGGCCCTTAACAAGAGAAAAAATAGAAGCACTAAGAGAAATCTGTGAA 2492

Qy    121 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 180
          || ||||||| || |||||| | || ||||| || || || ||||| || || ||||||
Db   2493 AAAATGGAAAGAGAAGGTCAGCTAGAAGAAGCGCCTCCAACTAATCCCTATAATACCCCT 2552

Qy    181 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 240
          |||||||| || |||||||||| || |||||||||| |||||||||||||||| |||
Db   2553 ACATTTGCAATTAAGAAAAAGGACAAAAACAAATGGAGGATGCTGATAGATTTTAGAGAA 2612

Qy    241 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 300
```


Db 2613 CTAAACAAGGTAACTCAAGATTTCACAGAGGTTCAGTTAGGAATTCCACACCCAGCAGGA 2672
Qy 301 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 360
Db 2673 TTAGCCAAGAAAAGAAGAATTACTGTGTTAGATGTAGGAGATGCCTACTTTTCCATACCC 2732
Qy 361 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 420
Db 2733 CTATATGAGGATTTTAGACAGTATACTGCATTTACTCTGCCATCAGTAAACAATGCAGAA 2792
Qy 421 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 480
Db 2793 CCAGGAAAAAGATATATATACAAAGTCTTACCACAGGGATGGAAGGGGTCACCAGCAATT 2852
Qy 481 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 540
Db 2853 TTTCAATACACAATGAGGCAAGTCTTAGAACCATTCAGAAAAGCAAACCCAGATGTCATT 2912
Qy 541 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 600
Db 2913 ATCGTTCAGTACATGGATGATATCTTAATAGCTAGTGACAGGACAGATTTGGAACATGAC 2972
Qy 601 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 660
Db 2973 AAAGTAGTCCTACAGCTAAAGGAACTTCTAAATGGCCTGGGATTTTCCACCCCCGACGAA 3032
Qy 661 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 720
Db 3033 AAGTTCCAAAAGGACCCTCCATACCAATGGATGGGCTATGAACTGTGGCCAACCAAATGG 3092
Qy 721 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 780
Db 3093 AAATTGCAAAAAATACAATTGCCCCAAAAGGAAGTATGGACAGTTAATGACATCCAGAAA 3152
Qy 781 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 840
Db 3153 CTAGTGGGTGTCCTAAACTGGGCGGCACAAATCTACCCAGGAATAAAGACCAAACACTTA 3212
Qy 841 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 900
Db 3213 TGTAAGCTAATTAGAGGAAAGATGACACCCACGGAAGAAGTACAGTGGACAGAATTAGCA 3272
Qy 901 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 960
Db 3273 GAAGCAGAGCTGGAGGAAAACAAAATTATCTTAAGCCAGGAACAGGAGGGACACTATTAC 3332
Qy 961 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 1020
Db 3333 CAAGAGGAAAAAGAGTTAGAAGCAACAGTTCAAAAGGATCAAGACAATCAGTGGACATAT 3392
Qy 1021 AAAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACA 1080
Db 3393 AAAGTACACCAGGGAGAAAAAATTCTAAAAGTAGGAAAATATGCAAAGATAAAAAATACC 3452
Qy 1081 CATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATA 1140
Db 3453 CATACCAACGGGGTCAGGTTGTTGGCACAGGTAGTTCAGAAAATAGGAAAAGAAGCACTA 3512
Qy 1141 GTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAG 1200
Db 3513 GTCATTTGGGGACGAATACCAAAATTTCACCTACCAGTAGAAAGAGAGACCTGGGAACAG 3572
Qy 1201 TGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCA 1260
Db 3573 TGGTGGGATAACTATTGGCAAGTGACATGGATCCCAGACTGGGACTTCGTATCCACCCCA 3632
Qy 1261 CCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTAT 1320
Db 3633 CCGTTGGTCAGGTTAGCATTTAACCTGGTAAAAGATCCTATACCAGGCGCAGAGACCTTC 3692
Qy 1321 TATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGAT 1380
Db 3693 TACACGGATGGATCTTGCAATAGGCAATCAAAAGAGGGAAAAGCAGGATATATAACAGAT 3752
Qy 1381 AGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAA 1440
Db 3753 AGAGGAAAAGACAAAGTAAGGATATTAGAGCAAACTACCAACCAACAAGCAGAATTAGAA 3812
Qy 1441 GCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAA 1500
Db 3813 GCCTTTGCAATGGCAGTAACAGACTCAGGTCCAAAAGTCAATATTGTAGTAGACTCACAG 3872
Qy 1501 TATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAA 1560
Db 3873 TATGTAATGGGAATAGTAACAGGCCAACCGGCTGAATCAGAGAGTAGAATAGTAAATAAA 3932
Qy 1561 ATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAA 1620
Db 3933 ATTATAGAAGAGATGATAAAAAAGGAAGCAATCTATGTTGCATGGGTCCCGGCCCACAAA 3992
Qy 1621 GGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTC 1680
Db 3993 GGCATAGGAGGAAATCAAGAAATTGACCACTTAGTAAGTCAGGGCATCAGACAAGTATTA 4052
Qy 1681 TTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAA 1740
Db 4053 TTCCTAGAGAGAATAGAGCCCGCTCAGGAAGAACATGGAAAATATCATAGCAATGTAAAA 4112
Qy 1741 GAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGT 1800
Db 4113 GAACTAGCCCATAAGTTTGGATTACCCAACCTGGTGGCAAGACAAATAGTAAACACATGT 4172
Qy 1801 GATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACT 1860
Db 4173 GCCCAGTGCCAACAAAAAGGGGAAGCTATACATGGGCAAGTAAATGCAGAACTAGGCACC 4232
Qy 1861 TGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCT 1920
Db 4233 TGGCAAATGGACTGCACACACTTAGAAGGAAAAATCATTATAGTAGCAGTACATGTTGCA 4292
Qy 1921 AGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTT 1980
Db 4293 AGTGGATTTATAGAAGCAGAAGTCATCCCACAGGAATCAGGAAGGCAAACAGCACTCTTC 4352
Qy 1981 CTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAAC 2040
Db 4353 CTATTAAAACTGGCCAGTAGGTGGCCAATAACACACTTGCACACAGATAATGGTGCCAAC 4412
Qy 2041 TTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGG 2100
Db 4413 TTCACTTCACAGGAGGTAAAGATGGTAGCATGGTGGGTAGGCATAGAACAATCCTTTGGA 4472
Qy 2101 GTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAAT 2160
Db 4473 GTACCTTACAATCCACAAAGCCAGGGAGTAGTAGAAGCAATGAATCACCACCTGAAAAAT 4532
Qy 2161 CAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTT 2220
Db 4533 CAGATAGAAAGAATTAGAGAGCAGGCAAATACAATGGAAACAATAGTACTAATGGCAGTT 4592
Qy 2221 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTA 2280
Db 4593 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACCCCAGTAGAAAGACTA 4652
Qy 2281 ATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTT 2340
Db 4653 GTCAATATGATCACCACAGAACAAGAAATACAATTCCTCCAAGCAAAAAATTCAAAATTA 4712
Qy 2341 AAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAG 2400
Db 4713 AAAAATTTTCGGGTCTATTTCAGAGAAGGCAGAAATCAACTGTGGCAAGGACCTGGGGAG 4772
Qy 2401 CTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTA 2460
Db 4773 CTACTGTGGAAAGGGGACGGAGCAGTCATAGTCAAGGTAGGGACAGATATAAAAGTAATA 4832

```
Qy   2461 CCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGT 2520
          || ||||||||||| || || |||| ||| ||||||  | |  ||||| |||||||| ||
Db   4833 CCAAGAAGAAAGGCCAAGATCATCAGAGACTATGGACCAAGGCAAGAGATGGATAGCGGT 4892

Qy   2521 TCCCACATGGAGGATACCGGAGAGG 2545
          |||||| |||||| | || | ||||
Db   4893 TCCCACCTGGAGGGTGCCAGGGAGG 4917


RESULT 13
AAN91774
ID   AAN91774 standard; DNA; 9822 BP.
XX
AC   AAN91774;
XX
DT   16-MAR-1990  (first entry)
XX
DE   Entire HIV-2/ST provirus DNA.
XX
KW   Antibodies
XX
OS   Human immunodeficiency virus-2.
XX
PN   WO8909815-A.
XX
PD   19-OCT-1989.
XX
PF   04-APR-1989;   89WO-US01379.
XX
PR   04-APR-1988;   88US-0177562.
XX
PA   (RESE ) RESEARCH CORPORATION TECHNOLOGIES INC.
XX
PI   Shaw GM,  Hahn BH,  Kong LI,  Lee SW;
XX
DR   WPI; 1989-324223/44.
XX
PT   Novel virus of the HIV-2 family - used in the prodn. of diagnostic
PT   reagents, vaccines and therapies for AIDS and AIDS-related diseases
XX
PS   Claim 37; page 47-53; 66pp; English.
XX
CC   This is used as a source of recombinant viral components in the
CC   development of diagnostic assays for HIV-2 viruses.
XX
SQ   Sequence 9822 BP; 3359 A; 2023 C; 2441 G; 1999 T; 0 other;

  Query Match           67.2%;  Score 1723;  DB 10;  Length 9822;
  Best Local Similarity   79.5%;  Pred. No. 0;
  Matches 2038;  Conservative    0;  Mismatches  525;  Indels    0;  Gaps    0;

Qy      2 TTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCAA 61
          | ||  | || ||  |||| ||  |||||  |     | ||||||||||||||||||||||
Db   2378 TACCAGTCGCCAAGATAGAACCAATAAAAATAATGCTGAAGCCAGGAAAGGATGGACCAA 2437

Qy     62 AATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAAA 121
          || |||  || ||||| ||| |||||||||| ||||  ||| ||| ||| |||||||| |
Db   2438 AACTGAGACAATGGCCCTTAACAAAAGAAAAAATAGAGGCACTAAAAGAGATCTGTGAGA 2497

Qy    122 AGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCCA 181
          | |||||||  || || ||| | ||||| || || || || ||||| || || |||||||
Db   2498 AAATGGAAAGAGAGGGCCAGCTAGAGGAGGCACCTCCAACTAATCCTTATAATACCCCCA 2557

Qy    182 CATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAAC 241
          ||||||| || |||||||||| || ||||||||||||||| ||||||||||| ||||
Db   2558 CATTTGCAATCAAGAAAAAGGACAAAAACAAATGGAGAATGCTAATAGATTTTAGAGAAC 2617

Qy    242 TAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGAC 301
          |||| | ||| ||||| ||||| || ||| |||| |||||||| |||||||| |||||||
Db   2618 TAAACAAGGTAACTCAAGACTTCACAGAAATCCAGTTAGGAATTCCACACCCAGCAGGAC 2677

Qy    302 TAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCTC 361
          |||| || | |||| |||||||| || || ||| |||| ||||| || || |||||||| |
Db   2678 TAGCCAAGAAGAAACGAATTACTGTCCTAGATGTAGGGGATGCTTACTTTTCCATACCAC 2737

Qy    362 TAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAGC 421
          || |||| || ||||| ||||| ||||| |||||| |||||||| |||| ||||| || |
Db   2738 TACATGAGGATTTTAGACAGTATACTGCATTTACTCTACCATCAATAAACAATGCTGAAC 2797

Qy    422 CAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATCT 481
          |||||||| ||||||| ||||| ||    || |||||||||||||| |||||||| || |
Db   2798 CAGGAAAAAGATACATATATAAAGTCTCACCACAGGGATGGAAGGGATCACCAGCAATTT 2857

Qy    482 TCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACCT 541
          | || ||||| ||||| || ||  ||||||| ||||| || ||||| || ||| | ||
Db   2858 TTCAGTACACAATGAGGCAGGTCTTAGAACCATTCAGAAAAGCAAACCCGGATATCATTC 2917

Qy    542 TAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGACA 601
          |  | ||||| |||||||| ||||| ||||| || |||||||||||  | ||||||||||
Db   2918 TCATTCAGTACATGGATGATATCTTGATAGCCAGCGACAGGACAGATTTAGAACATGACA 2977

Qy    602 GGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAGA 661
          | || ||| | |||  |||||||||  | ||| || | || ||||| ||||||||||||
Db   2978 GAGTGGTTCTGCAGCTAAAGGAACTTCTAAATGGCCTGGGATTTTCCACCCCAGATGAGA 3037

Qy    662 AATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGGA 721
          | |||||||||||  || ||||  |||||||||||  || ||| |||||||||| |||||||
Db   3038 AGTTCCAAAAAGACCCTCCATACCAATGGATGGGCTATGAACTGTGGCCAACTAAATGGA 3097

Qy    722 AGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAGT 781
          || ||||||  ||| | ||||| ||||  ||    |||||||| ||||| || ||||||
Db   3098 AGCTGCAAAGAATACAATTGCCCCAAAAGGAAGTATGGACAGTCAATGACATCCAAAAAC 3157

Qy    782 TAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAAACCAAACATCTCT 841
          | || || ||  |||||||||||||| ||||| || ||||| ||||| |||| | ||  | |
Db   3158 TGGTGGGTGTCCTAAATTGGGCAGCACAAATCTACCCAGGGATAAAGACCAGAAACTTAT 3217

Qy    842 GTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCAG 901
          |||||||||| ||||||||||||||| || ||||| || || |||||||| ||  | ||||
Db   3218 GTAGGTTAATCAGAGGAAAAATGACACTCACAGAAGAGGTACAGTGGACAGAATTAGCAG 3277

Qy    902 AAGCAGAATATGAGGAAAATAAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTACC 961
          |||| |||   || ||||| ||||| ||  | || ||||||||||||||||| |||||||
Db   3278 AAGCGGAACTAGAAGAAAACAAAATCATCTTAAGCCAGGAACAAGAAGGATGCTATTACC 3337

Qy    962 AAGAAGGCAAGCCATTAGAAGCCACCGTAATAAAGAGTCAGGACAATCAGTGGTCTTATA 1021
          |||| |  |||    ||||||| || ||   |||  ||| |||||||||||| | ||||
Db   3338 AAGAGGAAAAGGAGCTAGAAGCAACAGTCCAAAAAGATCAAGACAATCAGTGGACATATA 3397

Qy    1022 AAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACAC 1081
           | || ||||| | ||  ||||| || |||||||||||||| ||||||| |||| ||||| |
Db   3398 AGATACACCAGGGAGGAAAAATTCTAAAAGTAGGAAAATATGCAAAGGTAAAAAATACCC 3457

Qy    1082 ATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATAG 1141
           | ||||| ||||| ||||  ||||||| ||| | || |||||||||| |||||| |||
Db   3458 ACACCAACGGAGTCAGACTCCTAGCACAAGTAGTTCAAAAAATAGGAAAAGAAGCACTAG 3517

Qy    1142 TGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAGT 1201
           | || |||||||   | |||||||| ||| |||||||| || |  |||   ||||||||||
Db   3518 TCATTTGGGGACGAATACCAAAATTTCACCTACCAGTAGAAAGAGATACCTGGGAACAGT 3577

Qy    1202 GGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCAC 1261
           |||||    |||| ||||| || || ||||| || || ||||| || || || || || |
Db   3578 GGTGGGATAACTACTGGCAAGTGACATGGATCCCAGACTGGGACTTCATATCTACCCCGC 3637

Qy    1262 CGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTATT 1321
           | || || |||||||| || || || ||||| || || |||    || | |||||||  |
Db   3638 CACTGGTCAGATTAGTATTTAACCTGGTGAAAGATCCCATACTAGGCGCAGAAACCTTCT 3697

Qy    1322 ATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGATA 1381
           | |||||||||| || ||||| || |||| |||||| |||||||||| || |||||||
Db   3698 ACACAGATGGATCCTGCAATAAGCAATCAAGAGAAGGAAAAGCAGGATACATAACAGATA 3757

Qy    1382 GGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAAG 1441
           | || | |||||| || |   | ||||| || || || ||||| |||||||||| ||||
```

```
Db    3758 GAGGAAGAGACAAGGTGAGGCTATTAGAGCAAACCACCAATCAGCAAGCAGAATTAGAAG 3817

Qy    1442 CATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAAT 1501

Db    3818 CCTTTGCGATGGCAGTAACAGACTCAGGTCCAAAGGCCAACATTATAGTAGACTCACAAT 3877

Qy    1502 ATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAAA 1561

Db    3878 ATGTAATGGGAATAGTAGCAGGCCAACCAACAGAGTCAGAGAGTAAAATAGTAAATCAAA 3937

Qy    1562 TAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAAG 1621

Db    3938 TCATAGAAGAAATGATAAAAAAGGAAGCAATCTATGTTGCATGGGTCCCAGCCCATAAAG 3997

Qy    1622 GTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTCT 1681

Db    3998 GCATAGGAGGAAATCAGGAGGTAGATCACTTAGTAAGTCAGGGCATCAGACAAGTATTAT 4057

Qy    1682 TCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAAG 1741

Db    4058 TCCTAGAGAAAATAGAACCCGCTCAGGAGGAACATGAAAAATATCATAGCAATGTAAAAG 4117

Qy    1742 AATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGTG 1801

Db    4118 AACTATCCCATAAATTTGGACTGCCCAAATTAGTGGCAAGACAAATAGTAAACACATGCA 4177

Qy    1802 ATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACTT 1861

Db    4178 CCCAATGTCAGCAGAAAGGGGAGGCTATACATGGGCAAGTAAATGCAGAATTAGGCACTT 4237

Qy    1862 GGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCTA 1921

Db    4238 GGCAAATGGACTGCACACACTTAGAAGGAAAAATCATTATAGTAGCAGTACATGTTGCAA 4297

Qy    1922 GTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTTC 1981

Db    4298 GTGGATTTATAGAAGCAGAAGTCATCCCACAGGAATCAGGAAGGCAAACGGCACTCTTCC 4357

Qy    1982 TGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAACT 2041

Db    4358 TACTAAAACTGGCCAGTAGGTGGCCAATAACACATTTGCACACAGACAATGGTGCCAACT 4417

Qy    2042 TTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGGG 2101

Db    4418 TCACTTCACAGGAAGTAAAGATGGTGGCATGGTGGATAGGTATAGAACAATCCTTCGGAG 4477

Qy    2102 TACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAATC 2161

Db    4478 TACCTTACAATCCACAAAGCCAAGGAGTAGTGGAAGCAATGAATCACCACCTAAAAAATC 4537

Qy    2162 AAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTTC 2221

Db    4538 AGATAAGCAGAATTAGAGAGCAGGCAAACACAGTAGAAACAATAGTACTAATGGCAGTTC 4597

Qy    2222 ATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTAA 2281

Db    4598 ATTGCATGAATTTTAAAAGGAGGGGAGGAATAGGGGATATGACCCCAGCAGAAAGACTAA 4657

Qy    2282 TTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTTA 2341

Db    4658 TCAATATGGTCACTGCAGAACAGGAAATACAATTCCTCCAAGCAAAAAATTCAAAATTAC 4717

Qy    2342 AAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAGC 2401

Db    4718 AAAATTTTCGGGTCTATTTCAGAGAAGGCAGAGATCAGCTGTGGAAAGGACCTGGGGAAC 4777

Qy    2402 TATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTAC 2461

Db    4778 TACTGTGGAAGGGGGACGGAGCAGTCATAGTCAAGGTAGGGGCTGACATAAAAATAATAC 4837

Qy    2462 CCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGTT 2521

Db    4838 CAAGAAGGAAAGCTAAGATCATCAAAGACTATGGAGGAAGGCAAGAGATGGATAGCGGTT 4897
```

```
Qy    2522 CCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 2564

Db    4898 CCAACTTGGAGGGTGCCAGGGAGGATGGAGAGGTGGCATAGCC 4940
```

RESULT 14
AAZ89628
ID AAZ89628 standard; DNA; 9630 BP.
XX
AC AAZ89628;
XX
DT 11-JUL-2000 (first entry)
XX
DE HIV-2 ROD DNA.
XX
KW Diagnosis; probe; lymphocyte; virus; immunodiagnostic; infection;
KW antiviral; ss.
XX
OS Human immunodeficiency virus type 2.
XX
PN US6054565-A.
XX
PD 25-APR-2000.
XX
PF 28-APR-1994; 94US-0234875.
XX
PR 11-FEB-1987; 87US-0013477.
PR 03-SEP-1991; 91US-0752368.
PR 20-DEC-1991; 91US-0810908.
PR 03-MAR-1986; 86US-0835228.
PR 06-OCT-1986; 86US-0916080.
PR 21-NOV-1986; 86US-0933184.
PR 16-JAN-1987; 87US-0003764.
XX
PA (INSP) INST PASTEUR.
XX
PI Montagnier L, Clavel F, Guyader M, Geutard D, Sonigo P, Alizon M;
XX
DR WPI; 2000-328365/28.
XX
PT Novel cloned nucleotide sequences homologous or identical to the
PT portion of genomic RNA of HIV-2 viruses useful as probes and in
PT diagnostic tests to diagnose HIV-2 infection
XX
PS Claim 1; Column 10-23; 33pp; English.
XX
CC This invention describes a novel cloned nucleic acid (I) of a human
CC immunodeficiency virus type 2 (HIV-2). (I) is capable of being used
CC as probes in diagnostic method to obtain the immunological reagents
CC necessary to diagnose an HIV-2 infection. These sequences may be used
CC as probes in hybridization reactions with the genetic material of
CC infected patients to indicate whether the RNA of the HIV-2 virus is
CC present in these patient's lymphocytes or whether an analogous DNA is
CC present. The genetic sequence of the HIV-2 virus may be used to create
CC the polypeptides encoded by these sequences. Specifically, these
CC polypeptides may be created by expression of the cDNA obtained from
CC bacterial, yeast or animal cells. These polypeptides may be used in
CC diagnostic tests such as immunofluorescence assays, radioimmunoassays
CC (RIA) and Western Blot tests. Monoclonal antibodies to these polypeptides
CC of fragments may be created and used in immunodiagnostic tests. The
CC polypeptides of the present invention may also be used as immunogenic
CC reagents to induce protection against infection by HIV-2 viruses. The
CC polypeptides produced by recombinant-DNA techniques would function as
CC vaccine agents. The polypeptides may be used on competitive assays to
CC test the ability of various antiviral agents to determined their ability
CC to prevent the virus from fixing on its target. This sequence encodes
CC the HIV-2 ROD genome which is described in the method of the invention.
XX
SQ Sequence 9630 BP; 3310 A; 1959 C; 2373 G; 1976 T; 12 other;

Query Match 67.1%; Score 1721; DB 21; Length 9630;
Best Local Similarity 79.4%; Pred. No. 0;

Matches 2036; Conservative 1; Mismatches 526; Indels 0; Gaps 0;

```
Qy       2 TTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCAA 61
Db    2378 TACCAGTCGCCAAAGTAGAGCCAATAAAAATAATGCTAAAGCCAGGGAAAGATGGACCAA 2437

Qy      62 AATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAAA 121
Db    2438 AACTGAGACAATGGCCCTTAACAAAAGAAAAAATAGAAGCACTAAAAGAAATCTGTGAAA 2497

Qy     122 AGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCCA 181
Db    2498 AAATGGAAAAAGAAGGCCAGCTAGAGGAAGCACCTCCAACTAATCCTTATAATACCCCCA 2557

Qy     182 CATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAAC 241
Db    2558 CATTTGCAATCAAGAAAAAGGACAAAAACAAATGGAGGATGCTAATAGATTTCAGAGAAC 2617

Qy     242 TAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGAC 301
Db    2618 TAAACAAGGTAACTCAAGATTTCACAGAAATTCAGTTAGGAATTCCACACCCAGCAGGGT 2677

Qy     302 TAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCTC 361
Db    2678 TGGCCAAGAAGAGAAGAATTACTGTACTAGATGTAGGGGATGCTTACTTTTCCATACCAC 2737

Qy     362 TAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAGC 421
Db    2738 TACATGAGGACTTTAGACCATATACTGCATTTACTCTACCATCAGTGAACAATGCAGAAC 2797

Qy     422 CAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATCT 481
Db    2798 CAGGAAAAAGATACATATATAAAGTCTTGCCACAGGGATGGAAGGGATCACCAGCAATTT 2857

Qy     482 TCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACCT 541
Db    2858 TTCAACACACAATGAGACAGGTATTAGAACCATTCAGAAAAGCAAACAAGGATGTCATTA 2917

Qy     542 TAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGACA 601
Db    2918 TCATTCAGTACATGGATGATATCTTAATAGCTAGTGACAGGACAGATTTAGAACATGATA 2977

Qy     602 GGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAGA 661
Db    2978 GGGTAGTCCTGCAGCTCAAGGAACTTCTAAATGGCCTAGGATTTTCTACCCCAGATGAGA 3037

Qy     662 AATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGGA 721
Db    3038 AGTTCCAAAAAGACCCTCCATACCACTGGATGGGCTATGAACTATGGCCAACTAAATGGA 3097

Qy     722 AGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAGT 781
Db    3098 AGTTGCAGAAAATACAGTTGCCCCAAAAAGAAATATGGACAGTCAATGACATCCAGAAGC 3157

Qy     782 TAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTCT 841
Db    3158 TAGTGGGTGTCCTAAATTGGGCAGCACAACTCTACCCAGGGATAAAGACCAAACACTTAT 3217

Qy     842 GTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCAG 901
Db    3218 GTAGGTTAATCAGAGGAAAAATGACACTCACAGAAGAAGTACAGTGGACAGAATTAGCAG 3277

Qy     902 AAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTACC 961
Db    3278 AAGCAGAGCTAGAAGAAAACAGAATTATCCTAAHCCAHHAACAAGAGGGACACTATTACC 3337

Qy     962 AAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTATA 1021
Db    3338 AAGAAGAAAAAGAGCTAGAAGCAACAGTCCAAAAGGATCAAGAGAATCAGTGGACATATA 3397

Qy    1022 AAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACAC 1081
Db    3398 AAATACACCAGGAAGAAAAAATTCTAAAAGTAGGAAAATATGCAAAGGTGAAAAAGACCC 3457

Qy    1082 ATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATAG 1141
Db    3458 ATACCAATGGAATCAGATTGTTAGCACAGGTAGTTCAGAAAATAGGAAAAGAAGCACTAG 3517

Qy    1142 TGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAGT 1201
Db    3518 TCATTTGGGGACGAATACCAAAATTTCACCTACCAGTAGAGAGAGAAATCTCGGAGCAGT 3577

Qy    1202 GGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCAC 1261
Db    3578 GGTGGGATAACTACTGGCAAGTGACATGGATCCCAGACTGGGACTTCGTGTCTACCCCAC 3637

Qy    1262 CGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTATT 1321
Db    3638 CACTGGTCAGGTTAGCGTTTAACCTGGTAGGGGATCCTATACCAGGTGCAGAGACCTTCT 3697

Qy    1322 ATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGATA 1381
Db    3698 ACACAGATGGATCCTGCAATAGGCAATCAAAAGAAGGAAAAGCAGGATATGTAACAGATA 3757

Qy    1382 GGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAAG 1441
Db    3758 GAGGGAAAGACAAGGTAAAGAAACTAGAGCAAACTACCAATCAGCAAGCAGAACTAGAAG 3817

Qy    1442 CATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAAT 1501
Db    3818 CCTTTGCGATGGCACTAACAGACTCGGGTCCAAAAGTTAATATTATAGTAGACTCACAGT 3877

Qy    1502 ATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAAA 1561
Db    3878 ATGTAATGGGGATCAGTGCAAGCCAACCAACAGAGTCAGAAAGTAAAATAGTGAACCAGA 3937

Qy    1562 TAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAAG 1621
Db    3938 TCATAGAAGAAATGATAAAAAAGGAAGCAATCTATGTTGCATGGGTCCCAGCCCACAAAG 3997

Qy    1622 GTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTCT 1681
Db    3998 GCATAGGGGGAAACCAGGAAGTAGATCATTTAGTGAGTCAGGGTATCAGACAAGTGTTGT 4057

Qy    1682 TCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAAG 1741
Db    4058 TCCTGGAAAAAATAGAGCCCGCTCAGGAAGAACATGAAAAATATCATAGCAATGTAAAAG 4117

Qy    1742 AATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGTG 1801
Db    4118 AACTGTCTCATAAATTTGGAATACCCAATTTAGTGGCAAGGCAAATAGTAAACTCATGTG 4177

Qy    1802 ATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACTT 1861
Db    4178 CCCAATGTCAACAGAAAGGGGAAGCTATACATGGGCAAGTAAATGCAGAACTAGGCACTT 4237

Qy    1862 GGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCTA 1921
Db    4238 GGCAAATGGACTGCACACATTTAGAAGGAAAGATCATTATAGTAGCAGTACATGTTGCAA 4297

Qy    1922 GTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTTC 1981
Db    4298 GTGGATTTATAGAAGCAGAAGTCATCCCACAGGAATCAGGAAGACAAACAGCACTCTTCC 4357

Qy    1982 TGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAACT 2041
Db    4358 TATTGAAACTGGCAAGTAGGTGGCCAATAACACACTTGCATACAGATAATGGTGCCAACT 4417

Qy    2042 TTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGGG 2101
Db    4418 TCACTTCACAGGAGGTGAAGATGGTAGCATGGTGGATAGGTATAGAACAATCCTTTGGAG 4477

Qy    2102 TACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAATC 2161
Db    4478 TACCTTACAATCCACAGAGCCAAGGAGTAGTAGAAGCAATGAATCACCATCTAAAAAACC 4537
```

```
Qy    2162 AAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTTC 2221
           ||||||  |||||||||| ||||| |||||| || ||||||| |||||| |||||||| |||
Db    4538 AAATAAGTAGAATCAGAGAACAGGCAAATACAATAGAAACAATAGTACTAATGGCAATTC 4597

Qy    2222 ATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTAA 2281
           ||||||||||||||||||||||||| ||||||||||||||||||||| ||||||||||||
Db    4598 ATTGCATGAATTTTAAAAGAAGGGGGGGAATAGGGGATATGACTCCATCAGAAAGATTAA 4657

Qy    2282 TTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTTA 2341
           | || ||||||| |||||||||||||||||||| |  ||| | ||||| |||||||| |
Db    4658 TCAATATGATCACCACAGAACAAGAGATACAATTCCTCCAAGCCAAAAATTCAAAATTAA 4717

Qy    2342 AAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAGC 2401
           || ||||||||||||||| |||||||||||||||||||| ||||||| ||||| || || |
Db    4718 AAGATTTTCGGGTCTATTTCAGAGAAGGCAGAGATCAGTTGTGGAAAGGACCTGGGGAAC 4777

Qy    2402 TATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTAC 2461
           || |||||||||| |||||||||||| |  | |||||||| |||||||| ||  || |||
Db    4778 TACTGTGGAAAGGAGAAGGAGCAGTCCTAGTCAAGGTAGGAACAGACATAAAAATAATAC 4837

Qy    2462 CCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGTT 2521
           | ||||| || || || || |||| ||| |||||||||| || ||||| |||||||  |||
Db    4838 CAAGAAGGAAAGCCAAGATCATCAGAGACTATGGAGGAAGACAAGAGATGGATAGTGGTT 4897

Qy    2522 CCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 2564
           |||||:  |   | || | |||| | ||||  ||||||||||
Db    4898 CCCACCYHHAHHGTGCCAGGGAGGATGGAGAAATGGCATAGCC 4940
```

RESULT 15
AAQ02830
ID AAQ02830 standard; cDNA; 9360 BP.
XX
AC AAQ02830;
XX
DT 29-MAY-1989 (first entry)
XX
DE cDNA to HIV-2 RNA.
XX
KW HIV; AIDS; Vaccine; pUC-HIV-2(GH-1).
XX
OS HIV-2.
XX

FH	Key	Location/Qualifiers
FT	CDS	544..2109
FT		/*tag= a
FT		/label=Gag reading frame
FT	CDS	1830..4934
FT		/*tag= b
FT		/label=Pol gene reading frame.
FT	CDS	4867..5511
FT		/*tag= c
FT		/label=Q gene reading frame
FT	CDS	5342..5677
FT		/*tag= d
FT		/label=X gene reading frame
FT	CDS	5680..5994
FT		/*tag= e
FT		/label=R gene reading frame
FT	CDS	6148..8703
FT		/*tag= f
FT		/label=Env gene reading frame.
FT	CDS	8540..9304
FT		/*tag= g
FT		/label=F gene reading frame.

XX
PN JP01289486-A.
XX
PD 21-NOV-1989.
XX
PF 16-MAY-1988; 88JP-0119024.
XX
PR 16-MAY-1988; 88JP-0119024.
XX
PA (TOFU) TOA NENRYO KOGYO KK.
PA (FUJI-) FUJI KEBIO KK.
XX
DR WPI; 1990-005177/01.
DR P-PSDB; AAR04024-30.
XX
PT DNA indicating complement to RNA gene -
PT of Human Immunodeficiency Virus type 2 used for new vaccine or
PT diagnostic for AIDS virus.
XX
PS Claim 2; Fig.4; 12pp; Japanese.
XX
CC cDNA to novel HIV-2 (GH-1) has been integrated into plasmid
CC pUC HIV-2(GH-1). Useful for diagnosis and vaccination against the virrus.
CC Carries 7 overlapping genes in varrious reading frames, includding Gag,
CC Pol and Env.
XX
SQ Sequence 9360 BP; 3211 A; 1944 C; 2317 G; 1888 T; 0 other;

Query Match 67.0%; Score 1717.6; DB 11; Length 9360;
Best Local Similarity 79.4%; Pred. No. 0;
Matches 2035; Conservative 0; Mismatches 529; Indels 0; Gaps 0;

```
Qy       1 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 60
           || || || || ||  |||| ||  |||||||  | || |||||||| || |||||||||
Db    2376 TTACCAATCGCCAAGATAGAACCAATAAAAGTAACGTTGAAGCCAGGGAAAGATGGACCA 2435

Qy      61 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 120
           | | |||  || |||||  || ||||||| || ||||  ||| ||||||| ||||| |||
Db    2436 AGACTGAGACAATGGCCCCTAACAAAAGAGAAAATAGAAGCACTAAGAGAGATCTGCGAA 2495

Qy     121 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 180
           || |||||||||| || ||  | || || || || || || || ||||| || || ||||||
Db    2496 AAAATGGAAAAGGAGGGCCAACTAGAAGAGGCACCTCCAACTAATCCTTATAATACCCCC 2555

Qy     181 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 240
           |||||||| || |||||||||||| |||||||||||||| ||||| ||||| |||
Db    2556 ACATTTGCAATTAAGAAAAAGGACAAGAACAAATGGAGGATGCTAATAGACTTTAGAGAA 2615

Qy     241 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 300
           ||||| ||||| ||||| || || || ||| | ||  ||||||| |||||||| ||||||
Db    2616 CTAAACAGGGTGACTCAAGATTTCACAGAAATTCAGCTAGGAATTCCACACCCGGCAGGA 2675

Qy     301 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 360
           |||| |||| |||||| || || ||||| ||| |||| ||||| || || ||||||||
Db    2676 TTAGCCAAAAAGAAAAGGATCACTGTACTAGATGTAGGGGATGCCTACTTTTCCATACCG 2735

Qy     361 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 420
           ||| ||||||| |||||||||| ||||| |||||  |||||||||||||| ||||||||
Db    2736 CTACATGAAGATTTTAGGCAGTATACTGCATTTACCCTACCATCAGTAAACAATGCAGAA 2795

Qy     421 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 480
           ||||||||| |||| || |||||||  | || || |||||||||| |||||||| ||
Db    2796 CCAGGAAAAAGATATATATATAAGGTCTTACCACAAGGATGGAAGGGATCACCAGCAATT 2855

Qy     481 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 540
           || ||| |||| ||||| || ||  |||||||  ||||| || ||||| |||||||| |
Db    2856 TTTCAACACACAATGAGGCAAGTCTTAGAACCTTTCAGAAAAGCAAACCCAGATGTCATT 2915

Qy     541 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 600
           |  |||| || |||||||| || |||||||||||||||||||||    | || |||||||
Db    2916 CTCATCCAATACATGGATGATATTTTAATAGCTAGTGACAGGACAGGTTTAGAGCATGAC 2975

Qy     601 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 660
           |  |||||  | |||  ||| ||||||  ||||| || ||||||| |||||||||| |||
Db    2976 AAAGTAGTCCTGCAGCTAAAAGAACTTCTGAATGGCCTAGGGTTCTCTACCCCAGATGAG 3035

Qy     661 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 720
           || |||||||||| || ||  |||||||||||||| || ||| | |||||||| |||||||
Db    3036 AAGTTCCAAAAAGACCCTCCGCTTCAATGGATGGGCTATGAACTATGGCCAACTAAATGG 3095

Qy     721 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 780
```

```
Db   3096 AAGCTGCAGAAATTACAACTGCCCCAGAAAGAAATATGGACAGTCAACGACATCCAGAAG 3155

Qy    781 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 840

Db   3156 CTGGTGGGAGTCTTAAATTGGGCGGCACAAATCTATCCAGGAATAAAAACCAAACATTTG 3215

Qy    841 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 900

Db   3216 TGTAGGCTAATCAAAGGAAAAATGACACTCACAGAAGAAGTGCAGTGGACAGAGTTAGCA 3275

Qy    901 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 960

Db   3276 GAAGCAGAGCTAGAAGAAAACAAAATTATCTTAAGCCAGGAACAAGAAGGATACTATTAC 3335

Qy    961 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 1020

Db   3336 CAAGAAGAAAAAGAATTAGAGGCAACAATCCAAAAAAATCAAGACAATCAGTGGACATAC 3395

Qy   1021 AAAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACA 1080

Db   3396 AAAATACACCAGGAAGAGAAAATCCTAAAAGTAGGGAAATATGCAAAGATAAAAAATACC 3455

Qy   1081 CATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATA 1140

Db   3456 CATACCAATGGGGTCAGATTACTAGCACAGGTAGTTCAGAAAATAGGAAAAGAAGCACTA 3515

Qy   1141 GTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAG 1200

Db   3516 GTCATCTGGGGACGGATACCAAAATTCCACTTACCGGTGGAGAGAGAGACCTGGGAGCAG 3575

Qy   1201 TGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCA 1260

Db   3576 TGGTGGGATAACTACTGGCAAGTGACATGGATCCCAGAGTGGGACTTCGTATCTACCCCA 3635

Qy   1261 CCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTAT 1320

Db   3636 CCACTAGTCAGGTTAACATTTAACCTAGTAGGAGATCCTATACCAGGCGCAGAGACCTTC 3695

Qy   1321 TATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGAT 1380

Db   3696 TACACAGATGGATCATGCAATAGACAGTCAAAGGAAGGAAAAGCAGGATATGTGACAGAT 3755

Qy   1381 AGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAA 1440

Db   3756 AGAGGAAGAGACAAGGTAAGAGTACTAGAACGAACTACCAATCAGCAAGCAGAATTAGAA 3815

Qy   1441 GCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAA 1500

Db   3816 GCCTTTGCGATGGCACTGACAGACTCGGGTCCAAAAGTTAATATCATAGTAGACTCACAG 3875

Qy   1501 TATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAA 1560

Db   3876 TATGTAATGGGAATAGTAGTAGGCCAACCAACAGAGTCAGAAAGTAGAATAGTAAACCAA 3935

Qy   1561 ATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAA 1620

Db   3936 ATCATTGAGGACATGATAAAGAAAGAAGCAGTCTATGTCGCGTGGGTCCCAGCCCATAAA 3995

Qy   1621 GGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTC 1680

Db   3996 GGCATAGGAGGAAACCAGGAAGTAGACCATTTAGTAAGTCAGGGCATCAGACAAGTATTG 4055

Qy   1681 TTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAA 1740

Db   4056 TTCCTAGAAAGGATAGAGCCCGCCCAAGAAGAACATGAAAAATATCATAGCAATATGAAA 4115

Qy   1741 GAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGT 1800

Db   4116 GAACTAACCCATAAATTTGGAATACCCCAACTAGTAGCAAGACAGATAGTAAACACATGT 4175

Qy   1801 GATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACT 1860

Db   4176 GCCCAATGCCAACAGAAAGGAGAGGCCATACATGGGCAGGTAAATGCAGAAATAGGTGTC 4235

Qy   1861 TGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCT 1920

Db   4236 TGGCAAATGGACTGCACACACTTAGAAGGAAAGATCATTATAGTAGCAGTACATGTTGCA 4295

Qy   1921 AGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTT 1980

Db   4296 AGTGGATTCATAGAAGCAGAAGTTATCCCACAGGAATCAGGAAGGCAGACAGCGCTCTTC 4355

Qy   1981 CTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAAC 2040

Db   4356 CTATTAAAACTGGCCAGTAGGTGGCCAATAACACACTTACACACAGACAATGGCTCCAAC 4415

Qy   2041 TTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGG 2100

Db   4416 TTCACTTCACAGGAAGTGAAGATGGTGGCATGGTGGATAGGTATAGAGCAATCCTTTGGA 4475

Qy   2101 GTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAAT 2160

Db   4476 GTACCTTACAACCCACAAAGCCAGGGAGTAGTAGAAGCAATGAATCACCACTTAAAGAAT 4535

Qy   2161 CAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTT 2220

Db   4536 CAGATAAGTAGAATTAGAGAACAGGCAAATACAATAGAAACCATAGTACTAATGGCAGTT 4595

Qy   2221 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTA 2280

Db   4596 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACCCCAGCAGAAAGACTA 4655

Qy   2281 ATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTT 2340

Db   4656 ATCAACATGATTACCACAGAACAAGAAATACAATTCCTCCAAAGAAAAAATTCAAATTTT 4715

Qy   2341 AAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAG 2400

Db   4716 AAAAACTTCCAGGTCTATTACAGAGAAGGCAGAGATCAGCTGTGGAAAGGACCTGGAGAA 4775

Qy   2401 CTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTA 2460

Db   4776 CTACTGTGGAAGGGAGACGGAGCAGTCATAGTCAAGGTAGGGGCAGACATAAAAGTAATA 4835

Qy   2461 CCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGT 2520

Db   4836 CCAAGAAGGAAGGCCAAGATTATCAGAGACTATGGAGGAAGGCAAGAACTGGATAGTTCC 4895

Qy   2521 TCCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 2564

Db   4896 CACCTGGAGGGTGCCAGGGAGGAGGATGGAGAGGTGGCATAGTC 4939
```

Search completed: November 16, 2002, 07:56:17
Job time : 657.009 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 16, 2002, 06:34:38 ; Search time 59.4352 Seconds
(without alignments)
13229.850 Million cell updates/sec

Title: M33262_COPY_3105_5668
Perfect score: 2564
Sequence: 1 TTTCCCATAGCTAAAGTAGA.....GCTAGAGAGGTGGCATAGCC 2564

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

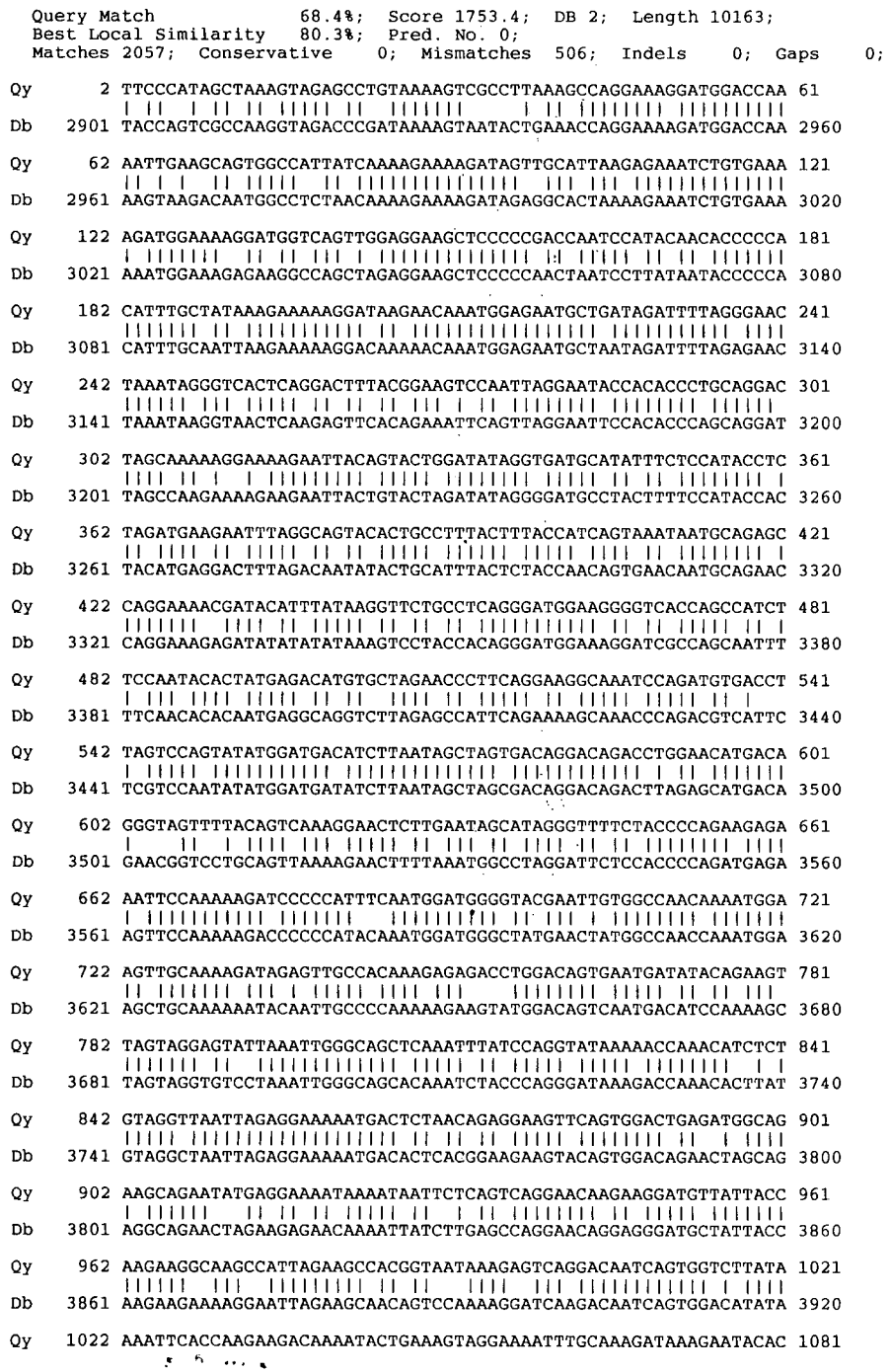
SUMMARIES

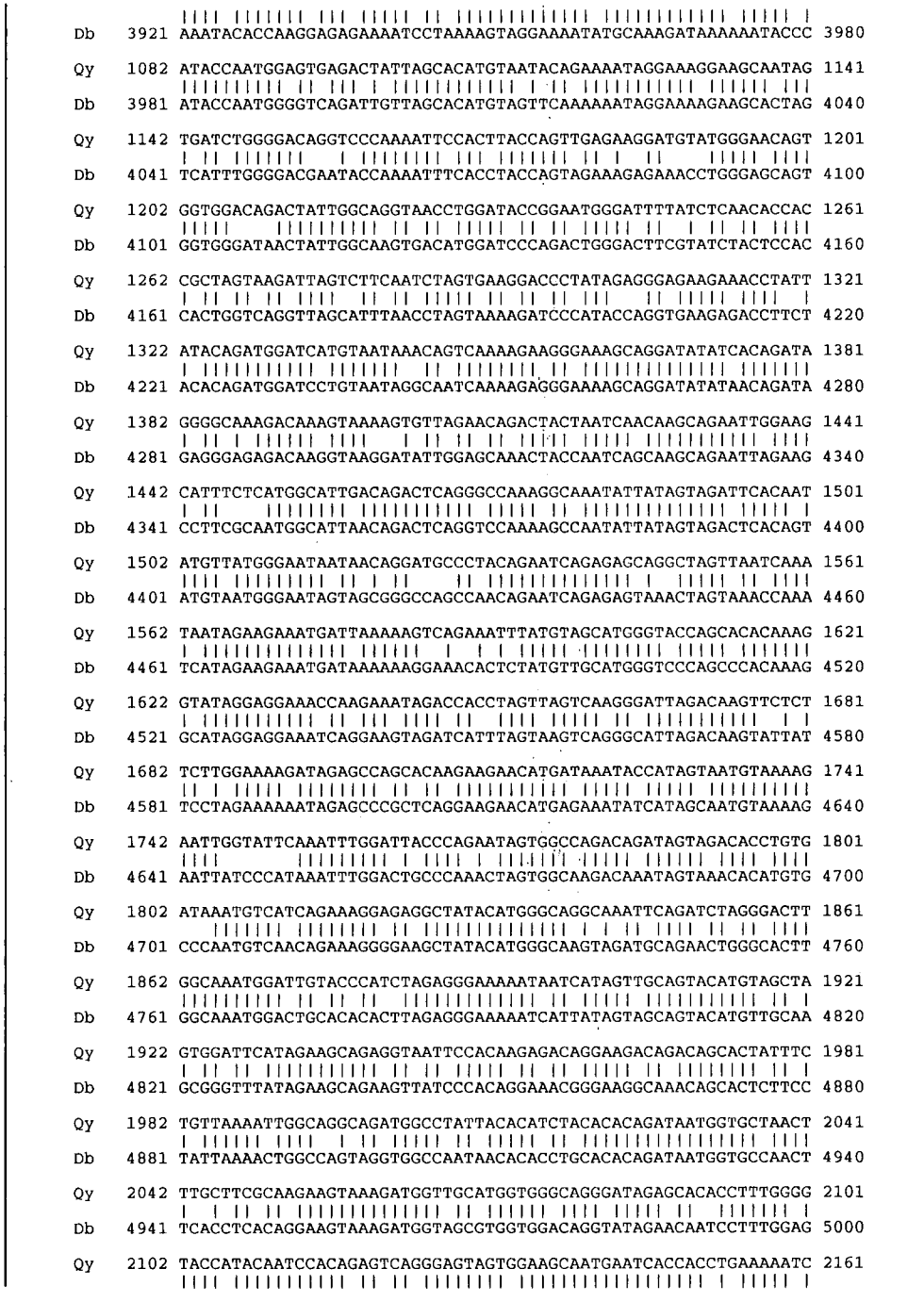
Result No.	Score	% Query Match	Length	DB	ID	Description
1	1753.4	68.4	10163	2	US-08-659-251-1	Sequence 1, Appli
2	1753.4	68.4	10163	4	US-09-256-490-1	Sequence 1, Appli
3	1753.4	68.4	10163	5	PCT-US96-11445-1	Sequence 1, Appli
4	1751.4	68.3	3168	2	US-08-659-251-14	Sequence 14, Appl
5	1751.4	68.3	3168	4	US-09-256-490-14	Sequence 14, Appl
6	1751.4	68.3	3168	5	PCT-US96-11445-14	Sequence 14, Appl
7	1738.4	67.8	9633	6	5223423-1	Patent No. 5223423
8	1587	61.9	2689	2	US-08-876-546A-15	Sequence 15, Appl
9	1587	61.9	2689	4	US-09-412-252-15	Sequence 15, Appl
10	1089.8	42.5	9746	1	US-08-022-835-3	Sequence 3, Appli
11	1089.8	42.5	9746	1	US-08-388-809-3	Sequence 3, Appli
12	1089.8	42.5	9746	2	US-08-647-714-3	Sequence 3, Appli
13	1083.4	42.3	5362	3	US-08-463-210-5	Sequence 5, Appli
14	1070.6	41.8	9737	2	US-08-944-449-7	Sequence 7, Appli
15	1070.6	41.8	9737	4	US-09-353-362-7	Sequence 7, Appli
16	1070.6	41.8	9739	1	US-08-022-835-1	Sequence 1, Appli
17	1070.6	41.8	9739	1	US-08-388-809-1	Sequence 1, Appli
18	1070.6	41.8	9739	2	US-08-647-714-1	Sequence 1, Appli
19	1070.2	41.7	9793	1	US-08-470-202-56	Sequence 56, Appl
20	1070.2	41.7	9793	1	US-08-471-770-56	Sequence 56, Appl
21	1070.2	41.7	9793	2	US-08-468-059-56	Sequence 56, Appl
22	1070.2	41.7	9793	4	US-09-109-916-56	Sequence 56, Appl
23	1069	41.7	7399	2	US-08-418-848A-9	Sequence 9, Appli
24	1069	41.7	9709	2	US-08-188-583-5	Sequence 5, Appli
25	1069	41.7	9709	3	US-08-388-353-1	Sequence 1, Appli
26	1069	41.7	9709	3	US-08-488-551B-1	Sequence 1, Appli
27	1069	41.7	9709	4	US-09-309-572-15	Sequence 15, Appl
28	1069	41.7	12494	4	US-08-935-312-13	Sequence 13, Appl
29	1069	41.7	12494	4	US-08-848-760B-33	Sequence 33, Appl
30	1069	41.7	15581	3	US-08-646-538-35	Sequence 35, Appl
31	1069	41.7	15581	4	US-09-503-222-35	Sequence 35, Appl
32	1067.4	41.6	2739	2	US-07-743-357-16	Sequence 16, Appl
33	1067.4	41.6	3033	2	US-07-743-357-19	Sequence 19, Appl
34	1067.4	41.6	3856	2	US-07-743-357-20	Sequence 20, Appl
35	1067.4	41.6	4113	2	US-07-743-357-21	Sequence 21, Appl
36	1067.4	41.6	8932	4	US-09-124-900-1	Sequence 1, Appli
37	1067.4	41.6	8933	3	US-08-463-210-4	Sequence 4, Appli
38	1067.4	41.6	8933	4	US-09-620-958A-3	Sequence 3, Appli
39	1062.6	41.4	8933	4	US-09-620-958A-4	Sequence 4, Appli
40	1056.2	41.2	8933	4	US-09-620-958A-9	Sequence 9, Appli
41	1053	41.1	9207	3	US-08-388-353-800	Sequence 800, App
42	1053	41.1	9207	3	US-08-488-551B-800	Sequence 800, App
43	958.4	37.4	2348	2	US-08-876-546A-1	Sequence 1, Appli
44	958.4	37.4	2348	4	US-09-412-252-1	Sequence 1, Appli
45	958.4	37.4	2348	4	US-09-079-675-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT 1
US-08-659-251-1
; Sequence 1, Application US/08659251
; Patent No. 5883081
;  GENERAL INFORMATION:
;    APPLICANT:  Kraus, Guenter
;    APPLICANT:  Wong-Staal, Flossie
;    APPLICANT:  Talbott, Randy
;    APPLICANT:  Poeschla, Eric
;    TITLE OF INVENTION:  Isolation of No. 5883081el HIV-2 Proviruses
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew LLP
;      STREET:  Two Embarcadero Center, Eighth Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/659,251
;      FILING DATE:  No. 5883081 yet assigned
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/001,441
;      FILING DATE:  26-JUL-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Garrett-Wackowski, Eugenia
;      REGISTRATION NUMBER:  37,330
;      REFERENCE/DOCKET NUMBER:  02307E-056410US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 576-0200
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10163 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  -
;      LOCATION:  1..10163
;      OTHER INFORMATION:  /note= "HIV-2KR provirus"
US-08-659-251-1
```

Query Match 68.4%; Score 1753.4; DB 2; Length 10163;
Best Local Similarity 80.3%; Pred. No. 0;
Matches 2057; Conservative 0; Mismatches 506; Indels 0; Gaps 0;





```
Db    5001 TACCTTACAATCCACAAAGCCAAGGAGTAGTAGAAGCAATGAATCACCACTTAAAAAACC 5060

Qy    2162 AAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTTC 2221
           | |||   ||||| || || || |||||| || | ||||| ||||||||||||||||||
Db    5061 AGATAAGCAGAATTAGAGAGCAGGCAAATACAATGGAAACAATAGTATTAATGGCAGTTC 5120

Qy    2222 ATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTAA 2281
           ||||||||||||||||||||||||||||||||||||||||||||| |||||||||||| |||
Db    5121 ATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACCCCAGCAGAAAGACTAA 5180

Qy    2282 TTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTTA 2341
           | || |||||||| |||||||||| |||||||| |  ||  ||||||| |||||||| |
Db    5181 TCAATATGATCACCACAGAACAAGAAATACAATTCCTCCACGCAAAAAATTCAAAATTAA 5240

Qy    2342 AAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAGC 2401
           ||||||| |||||||||| |||||||||||||||||| |||||||| ||||| || || |
Db    5241 AAAATTTCCGGGTCTATTTCAGAGAAGGCAGAGATCAGCTGTGGAAAGGACCTGGGGAAC 5300

Qy    2402 TATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTAC 2461
           || ||||||| || || |||||||||||  | ||||||||||||||||| ||  |||| |
Db    5301 TACTGTGGAAGGGAGATGGAGCAGTCATAGTCAAGGTAGGGACAGACATAAAAATAGTGC 5360

Qy    2462 CCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGTT 2521
           | ||||| || ||||| || |||| ||| |||||||| |   ||||||||||| ||||
Db    5361 CAAGAAGGAAAGCTAAGATCATCAGAGACTATGGAGGAAGGCGAGAGGTGGATAGTAGTT 5420

Qy    2522 CCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 2564
           ||||| |||||| |||| | |||| | |||| ||||||||||
Db    5421 CCCACTTGGAGGGTACCAGGGAGGATGGAGAAGTGGCATAGCC 5463


RESULT 2
US-09-256-490-1
; Sequence 1, Application US/09256490
; Patent No. 6235881
;  GENERAL INFORMATION:
;    APPLICANT:  Kraus, Guenter
;    APPLICANT:  Wong-Staal, Flossie
;    APPLICANT:  Talbott, Randy
;    APPLICANT:  Poeschla, Eric
;    TITLE OF INVENTION:  Isolation of No. 6235881el HIV-2 Proviruses
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew LLP
;      STREET:  Two Embarcadero Center, Eighth Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/256,490
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/659,251
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Garrett-Wackowski, Eugenia
;      REGISTRATION NUMBER:  37,330
;      REFERENCE/DOCKET NUMBER:  02307E-056410US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 576-0200
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10163 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  -
;      LOCATION:  1..10163
;      OTHER INFORMATION:  /note= "HIV-2KR provirus"
US-09-256-490-1

  Query Match             68.4%;  Score 1753.4;  DB 4;  Length 10163;
  Best Local Similarity   80.3%;  Pred. No. 0;
  Matches 2057;  Conservative    0;  Mismatches  506;  Indels    0;  Gaps    0;

Qy       2 TTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCAA 61
           | ||  | || || ||||| ||  |||||||     | || |||||||| ||||||||||
Db    2901 TACCAGTCGCCAAGGTAGACCCGATAAAAGTAATACTGAAACCAGGAAAAGATGGACCAA 2960

Qy      62 AATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAAA 121
           || | |  || |||||  || ||||||||||||||  ||| ||| ||||||||||||||||
Db    2961 AAGTAAGACAATGGCCTCTAACAAAAGAAAAGATAGAGGCACTAAAAGAAATCTGTGAAA 3020

Qy     122 AGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCCA 181
           | |||||||  || || ||| | |||||||||||||| || ||||| || || ||||||||
Db    3021 AAATGGAAAGAGAAGGCCAGCTAGAGGAAGCTCCCCCAACTAATCCTTATAATACCCCCA 3080

Qy     182 CATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAAC 241
           ||||||| || |||||||||| || |||||||||||||||| |||||||||| ||||| ||||
Db    3081 CATTTGCAATTAAGAAAAAGGACAAAAACAAATGGAGAATGCTAATAGATTTTAGAGAAC 3140

Qy     242 TAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGAC 301
           |||||| ||| ||||| || || || ||| | || |||||||| |||||||| ||||||
Db    3141 TAAATAAGGTAACTCAAGAGTTCACAGAAATTCAGTTAGGAATTCCACACCCAGCAGGAT 3200

Qy     302 TAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCTC 361
           |||| || |  | |||||||||| ||||| |||||||| ||||| || || |||||||| |
Db    3201 TAGCCAAGAAAAGAAGAATTACTGTACTAGATATAGGGGATGCCTACTTTTCCATACCAC 3260

Qy     362 TAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAGC 421
           || |||| || ||||| || || ||||| ||||||  ||||| |||| || |||||||| |
Db    3261 TACATGAGGACTTTAGACAATATACTGCATTTACTCTACCAACAGTGAACAATGCAGAAC 3320

Qy     422 CAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATCT 481
           |||||||  |||| || ||||| || || || |||||||||| || || ||||| || |
Db    3321 CAGGAAAGAGATATATATATAAAGTCCTACCACAGGGATGGAAAGGATCGCCAGCAATTT 3380

Qy     482 TCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACCT 541
           | ||| |||| ||||| || ||  |||| || ||||| || ||||| ||||| || |
Db    3381 TTCAACACACAATGAGGCAGGTCTTAGAGCCATTCAGAAAAGCAAACCCAGACGTCATTC 3440

Qy     542 TAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGACA 601
           | ||||| |||||||||||| |||||||||||||| |||||||||||| | || ||||||||
Db    3441 TCGTCCAATATATGGATGATATCTTAATAGCTAGCGACAGGACAGACTTAGAGCATGACA 3500

Qy     602 GGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAGA 661
           |    ||  | |||| ||| ||||| || ||| || |||| || || ||||||| ||||
Db    3501 GAACGGTCCTGCAGTTAAAAGAACTTTTAAATGGCCTAGGATTCTCCACCCCAGATGAGA 3560

Qy     662 AATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGGA 721
           | |||||||||||| |||||||   |||||||||| || ||| | |||||||| |||||||
Db    3561 AGTTCCAAAAAGACCCCCCATACAAATGGATGGGCTATGAACTATGGCCAACCAAATGGA 3620

Qy     722 AGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAGT 781
           || ||||||| ||| | ||||| |||| ||| |||||||| ||||| || || |||
Db    3621 AGCTGCAAAAAATACAATTGCCCCAAAAAGAAGTATGGACAGTCAATGACATCCAAAAGC 3680

Qy     782 TAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTCT 841
           ||||||| ||  |||||||||||| ||||| || ||||| ||||| |||||||  | |
Db    3681 TAGTAGGTGTCCTAAATTGGGCAGCACAAATCTACCCAGGGATAAAGACCAAACACTTAT 3740

Qy     842 GTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCAG 901
           ||||| |||||||||||||||||| || || || ||||| |||||||| || || ||||
Db    3741 GTAGGCTAATTAGAGGAAAAATGACACTCACGGAAGAAGTACAGTGGACAGAACTAGCAG 3800

Qy     902 AAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTACC 961
```

```
Db   3801 AGGCAGAACTAGAAGAGAACAAAATTATCTTGAGCCAGGAACAGGAGGGATGCTATTACC 3860

Qy    962 AAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTATA 1021

Db   3861 AAGAAGAAAAGGAATTAGAAGCAACAGTCCAAAAGGATCAAGACAATCAGTGGACATATA 3920

Qy   1022 AAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACAC 1081

Db   3921 AAATACACCAAGGAGAGAAAATCCTAAAAGTAGGAAAATATGCAAAGATAAAAAATACCC 3980

Qy   1082 ATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATAG 1141

Db   3981 ATACCAATGGGGTCAGATTGTTAGCACATGTAGTTCAAAAAATAGGAAAAGAAGCACTAG 4040

Qy   1142 TGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAGT 1201

Db   4041 TCATTTGGGGACGAATACCAAAATTTCACCTACCAGTAGAAAGAGAAACCTGGGAGCAGT 4100

Qy   1202 GGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCAC 1261

Db   4101 GGTGGGATAACTATTGGCAAGTGACATGGATCCCAGACTGGGACTTCGTATCTACTCCAC 4160

Qy   1262 CGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTATT 1321

Db   4161 CACTGGTCAGGTTAGCATTTAACCTAGTAAAAGATCCCATACCAGGTGAAGAGACCTTCT 4220

Qy   1322 ATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGATA 1381

Db   4221 ACACAGATGGATCCTGTAATAGGCAATCAAAAGAGGGAAAAGCAGGATATATAACAGATA 4280

Qy   1382 GGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAAG 1441

Db   4281 GAGGGAGAGACAAGGTAAGGATATTGGAGCAAACTACCAATCAGCAAGCAGAATTAGAAG 4340

Qy   1442 CATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAAT 1501

Db   4341 CCTTCGCAATGGCATTAACAGACTCAGGTCCAAAAGCCAATATTATAGTAGACTCACAGT 4400

Qy   1502 ATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAAA 1561

Db   4401 ATGTAATGGGAATAGTAGCGGGCCAGCCAACAGAATCAGAGAGTAAACTAGTAAACCAAA 4460

Qy   1562 TAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAAG 1621

Db   4461 TCATAGAAGAAATGATAAAAAAGGAAACACTCTATGTTGCATGGGTCCCAGCCCACAAAG 4520

Qy   1622 GTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTCT 1681

Db   4521 GCATAGGAGGAAATCAGGAAGTAGATCATTTAGTAAGTCAGGGCATTAGACAAGTATTAT 4580

Qy   1682 TCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAAG 1741

Db   4581 TCCTAGAAAAAATAGAGCCCGCTCAGGAAGAACATGAGAAATATCATAGCAATGTAAAAG 4640

Qy   1742 AATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGTG 1801

Db   4641 AATTATCCCATAAATTTGGACTGCCCAAACTAGTGGCAAGACAAATAGTAAACACATGTG 4700

Qy   1802 ATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACTT 1861

Db   4701 CCCAATGTCAACAGAAAGGGGAAGCTATACATGGGCAAGTAGATGCAGAACTGGGCACTT 4760

Qy   1862 GGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCTA 1921

Db   4761 GGCAAATGGACTGCACACACTTAGAGGGAAAAATCATTATAGTAGCAGTACATGTTGCAA 4820

Qy   1922 GTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTTC 1981

Db   4821 GCGGGTTTATAGAAGCAGAAGTTATCCCACAGGAAACGGGAAGGCAAACAGCACTCTTCC 4880

Qy   1982 TGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAACT 2041

Db   4881 TATTAAAACTGGCCAGTAGGTGGCCAATAACACACCTGCACACAGATAATGGTGCCAACT 4940

Qy   2042 TTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGGG 2101

Db   4941 TCACCTCACAGGAAGTAAAGATGGTAGCGTGGTGGACAGGTATAGAACAATCCTTTGGAG 5000

Qy   2102 TACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAATC 2161

Db   5001 TACCTTACAATCCACAAAGCCAAGGAGTAGTAGAAGCAATGAATCACCACTTAAAAAACC 5060

Qy   2162 AAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTTC 2221

Db   5061 AGATAAGCAGAATTAGAGAGCAGGCAAATACAATGGAAACAATAGTATTAATGGCAGTTC 5120

Qy   2222 ATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTAA 2281

Db   5121 ATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACCCCAGCAGAAAGACTAA 5180

Qy   2282 TTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTTA 2341

Db   5181 TCAATATGATCACCACAGAACAAGAAATACAATTCCTCCACGCAAAAAATTCAAAATTAA 5240

Qy   2342 AAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAGC 2401

Db   5241 AAAATTTCCGGGTCTATTTCAGAGAAGGCAGAGATCAGCTGTGGAAAGGACCTGGGGAAC 5300

Qy   2402 TATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTAC 2461

Db   5301 TACTGTGGAAGGGAGATGGAGCAGTCATAGTCAAGGTAGGGACAGACATAAAAATAGTGC 5360

Qy   2462 CCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGTT 2521

Db   5361 CAAGAAGGAAAGCTAAGATCATCAGAGACTATGGAGGAAGGCGAGAGGTGGATAGTAGTT 5420

Qy   2522 CCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 2564

Db   5421 CCCACTTGGAGGGTACCAGGGAGGATGGAGAAGTGGCATAGCC 5463
```

RESULT 3
PCT-US96-11445-1
; Sequence 1, Application PC/TUS9611445
; GENERAL INFORMATION:
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Isolation of Novel HIV-2 Proviruses
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Robbins, Berliner & Carson
; STREET: 201 N. Figueroa Street, 5th Floor
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90012-2628
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/11445
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Berliner, Robert
; REGISTRATION NUMBER: 20,121
; REFERENCE/DOCKET NUMBER: 5555-399C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 977-1001
; TELEFAX: (213) 977-1003
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10163 base pairs
; TYPE: nucleic acid

```
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  -
;       LOCATION:  1..10163
;       OTHER INFORMATION:  /note= "HIV-2KR provirus"
PCT-US96-11445-1

  Query Match             68.4%;  Score 1753.4;  DB 5;  Length 10163;
  Best Local Similarity   80.3%;  Pred. No. 0;
  Matches 2057;  Conservative    0;  Mismatches  506;  Indels    0;  Gaps    0;

Qy       2 TTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCAA 61
Db    2901 TACCAGTCGCCAAGGTAGACCCGATAAAAGTAATACTGAAACCAGGAAAAGATGGACCAA 2960

Qy      62 AATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAAA 121
Db    2961 AAGTAAGACAATGGCCTCTAACAAAAGAAAAGATAGAGGCACTAAAAGAAATCTGTGAAA 3020

Qy     122 AGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCCA 181
Db    3021 AAATGGAAAGAGAAGGCCAGCTAGAGGAAGCTCCCCCAACTAATCCTTATAATACCCCCA 3080

Qy     182 CATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAAC 241
Db    3081 CATTTGCAATTAAGAAAAAGGACAAAAACAAATGGAGAATGCTAATAGATTTTAGAGAAC 3140

Qy     242 TAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGAC 301
Db    3141 TAAATAAGGTAACTCAAGAGTTCACAGAAATTCAGTTAGGAATTCCACACCCAGCAGGAT 3200

Qy     302 TAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCTC 361
Db    3201 TAGCCAAGAAAAGAAGAATTACTGTACTAGATATAGGGGATGCCTACTTTTCCATACCAC 3260

Qy     362 TAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAGC 421
Db    3261 TACATGAGGACTTTAGACAATATACTGCATTTACTCTACCAACAGTGAACAATGCAGAAC 3320

Qy     422 CAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATCT 481
Db    3321 CAGGAAAGAGATATATATATAAAGTCCTACCACAGGGATGGAAAGGATCGCCAGCAATTT 3380

Qy     482 TCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACCT 541
Db    3381 TTCAACACACAATGAGGCAGGTCTTAGAGCCATTCAGAAAAGCAAACCCAGACGTCATTC 3440

Qy     542 TAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGACA 601
Db    3441 TCGTCCAATATATGGATGATATCTTAATAGCTAGCGACAGGACAGACTTAGAGCATGACA 3500

Qy     602 GGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAGA 661
Db    3501 GAACGGTCCTGCAGTTAAAAGAACTTTTAAATGGCCTAGGATTCTCCACCCCAGATGAGA 3560

Qy     662 AATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGGA 721
Db    3561 AGTTCCAAAAAGACCCCCCATACAAATGGATGGGCTATGAACTATGGCCAACCAAATGGA 3620

Qy     722 AGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAGT 781
Db    3621 AGCTGCAAAAAATACAATTGCCCCAAAAAGAAGTATGGACAGTCAATGACATCCAAAAGC 3680

Qy     782 TAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTCT 841
Db    3681 TAGTAGGTGTCCTAAATTGGGCAGCACAAATCTACCCAGGGATAAAGACCAAACACTTAT 3740

Qy     842 GTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCAG 901
Db    3741 GTAGGCTAATTAGAGGAAAAATGACACTCACGGAAGAAGTACAGTGGACAGAACTAGCAG 3800

Qy     902 AAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTACC 961
Db    3801 AGGCAGAACTAGAAGAGAACAAAATTATCTTGAGCCAGGAACAGGAGGGATGCTATTACC 3860

Qy     962 AAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTATA 1021
Db    3861 AAGAAGAAAAGGAATTAGAAGCAACAGTCCAAAAGGATCAAGACAATCAGTGGACATATA 3920

Qy    1022 AAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACAC 1081
Db    3921 AAATACACCAAGGAGAGAAAATCCTAAAAGTAGGAAAATATGCAAAGATAAAAAATACCC 3980

Qy    1082 ATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATAG 1141
Db    3981 ATACCAATGGGGTCAGATTGTTAGCACATGTAGTTCAAAAAATAGGAAAAGAAGCACTAG 4040

Qy    1142 TGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAGT 1201
Db    4041 TCATTTGGGGACGAATACCAAAATTTCACCTACCAGTAGAAAGAGAAACCTGGGAGCAGT 4100

Qy    1202 GGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCAC 1261
Db    4101 GGTGGGATAACTATTGGCAAGTGACATGGATCCCAGACTGGGACTTCGTATCTACTCCAC 4160

Qy    1262 CGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTATT 1321
Db    4161 CACTGGTCAGGTTAGCATTTAACCTAGTAAAAGATCCCATACCAGGTGAAGAGACCTTCT 4220

Qy    1322 ATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGATA 1381
Db    4221 ACACAGATGGATCCTGTAATAGGCAATCAAAAGAGGGAAAAGCAGGATATATAACAGATA 4280

Qy    1382 GGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAAG 1441
Db    4281 GAGGGAGAGACAAGGTAAGGATATTGGAGCAAACTACCAATCAGCAAGCAGAATTAGAAG 4340

Qy    1442 CATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAAT 1501
Db    4341 CCTTCGCAATGGCATTAACAGACTCAGGTCCAAAAGCCAATATTATAGTAGACTCACAGT 4400

Qy    1502 ATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAAA 1561
Db    4401 ATGTAATGGGAATAGTAGCGGGCCAGCCAACAGAATCAGAGAGTAAACTAGTAAACCAAA 4460

Qy    1562 TAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAAG 1621
Db    4461 TCATAGAAGAAATGATAAAAAAGGAAACACTCTATGTTGCATGGGTCCCAGCCCACAAAG 4520

Qy    1622 GTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTCT 1681
Db    4521 GCATAGGAGGAAATCAGGAAGTAGATCATTTAGTAAGTCAGGGCATTAGACAAGTATTAT 4580

Qy    1682 TCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAAG 1741
Db    4581 TCCTAGAAAAAATAGAGCCCGCTCAGGAAGAACATGAGAAATATCATAGCAATGTAAAAG 4640

Qy    1742 AATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGTG 1801
Db    4641 AATTATCCCATAAATTTGGACTGCCCAAACTAGTGGCAAGACAAATAGTAAACACATGTG 4700

Qy    1802 ATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACTT 1861
Db    4701 CCCAATGTCAACAGAAAGGGGAAGCTATACATGGGCAAGTAGATGCAGAACTGGGCACTT 4760

Qy    1862 GGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCTA 1921
Db    4761 GGCAAATGGACTGCACACACTTAGAGGGAAAAATCATTATAGTAGCAGTACATGTTGCAA 4820

Qy    1922 GTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTTC 1981
Db    4821 GCGGGTTTATAGAAGCAGAAGTTATCCCACAGGAAACGGGAAGGCAAACAGCACTCTTCC 4880

Qy    1982 TGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAACT 2041
```

```
Db    4881 TATTAAAACTGGCCAGTAGGTGGCCAATAACACACCTGCACACAGATAATGGTGCCAACT 4940

Qy    2042 TTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGGG 2101

Db    4941 TCACCTCACAGGAAGTAAAGATGGTAGCGTGGTGGACAGGTATAGAACAATCCTTTGGAG 5000

Qy    2102 TACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAATC 2161

Db    5001 TACCTTACAATCCACAAAGCCAAGGAGTAGTAGAAGCAATGAATCACCACTTAAAAAACC 5060

Qy    2162 AAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTTC 2221

Db    5061 AGATAAGCAGAATTAGAGAGCAGGCAAATACAATGGAAACAATAGTATTAATGGCAGTTC 5120

Qy    2222 ATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTAA 2281

Db    5121 ATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACCCCAGCAGAAAGACTAA 5180

Qy    2282 TTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTTA 2341

Db    5181 TCAATATGATCACCACAGAACAAGAAATACAATTCCTCCACGCAAAAAATTCAAAATTAA 5240

Qy    2342 AAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAGC 2401

Db    5241 AAAATTTCCGGGTCTATTTCAGAGAAGGCAGAGATCAGCTGTGGAAAGGACCTGGGGAAC 5300

Qy    2402 TATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTAC 2461

Db    5301 TACTGTGGAAGGGAGATGGAGCAGTCATAGTCAAGGTAGGGACAGACATAAAAATAGTGC 5360

Qy    2462 CCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGTT 2521

Db    5361 CAAGAAGGAAAGCTAAGATCATCAGAGACTATGGAGGAAGGCGAGAGGTGGATAGTAGTT 5420

Qy    2522 CCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 2564

Db    5421 CCCACTTGGAGGGTACCAGGGAGGATGGAGAAGTGGCATAGCC 5463


RESULT 4
US-08-659-251-14
; Sequence 14, Application US/08659251
; Patent No. 5883081
;  GENERAL INFORMATION:
;    APPLICANT:  Kraus, Guenter
;    APPLICANT:  Wong-Staal, Flossie
;    APPLICANT:  Talbott, Randy
;    APPLICANT:  Poeschla, Eric
;    TITLE OF INVENTION:  Isolation of No. 5883081el HIV-2 Proviruses
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew LLP
;      STREET:  Two Embarcadero Center, Eighth Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/659,251
;      FILING DATE:  No. 5883081 yet assigned
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/001,441
;      FILING DATE:  26-JUL-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Garrett-Wackowski, Eugenia
;      REGISTRATION NUMBER:  37,330
;      REFERENCE/DOCKET NUMBER:  02307E-056410US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 576-0200
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3168 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    FEATURE:
;      NAME/KEY:  -
;      LOCATION:  1..3168
;      OTHER INFORMATION:  /label= pol
;      OTHER INFORMATION:  /note= "HIV-2KR subsequence encoding
;      OTHER INFORMATION:  pol gene"
US-08-659-251-14

  Query Match             68.3%;  Score 1751.4;  DB 2;  Length 3168;
  Best Local Similarity   80.2%;  Pred. No. 0;
  Matches 2055;  Conservative    0;  Mismatches  506;  Indels    0;  Gaps    0;

Qy       2 TTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCAA 61

Db     608 TACCAGTCGCCAAGGTAGACCCGATAAAAGTAATACTGAAACCAGGAAAAGATGGACCAA 667

Qy      62 AATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAAA 121

Db     668 AAGTAAGACAATGGCCTCTAACAAAAGAAAAGATAGAGGCACTAAAAGAAATCTGTGAAA 727

Qy     122 AGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCCA 181

Db     728 AAATGGAAAGAGAAGGCCAGCTAGAGGAAGCTCCCCCAACTAATCCTTATAATACCCCCA 787

Qy     182 CATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAAC 241

Db     788 CATTTGCAATTAAGAAAAAGGACAAAAACAAATGGAGAATGCTAATAGATTTTAGAGAAC 847

Qy     242 TAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGAC 301

Db     848 TAAATAAGGTAACTCAAGAGTTCACAGAAATTCAGTTAGGAATTCCACACCCAGCAGGAT 907

Qy     302 TAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCTC 361

Db     908 TAGCCAAGAAAAGAAGAATTACTGTACTAGATATAGGGGATGCCTACTTTTCCATACCAC 967

Qy     362 TAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAGC 421

Db     968 TACATGAGGACTTTAGACAATATACTGCATTTACTCTACCAACAGTGAACAATGCAGAAC 1027

Qy     422 CAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATCT 481

Db    1028 CAGGAAAGAGATATATATATAAAGTCCTACCACAGGGATGGAAAGGATCGCCAGCAATTT 1087

Qy     482 TCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACCT 541

Db    1088 TTCAACACACAATGAGGCAGGTCTTAGAGCCATTCAGAAAAGCAAACCCAGACGTCATTC 1147

Qy     542 TAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGACA 601

Db    1148 TCGTCCAATATATGGATGATATCTTAATAGCTAGCGACAGGACAGACTTAGAGCATGACA 1207

Qy     602 GGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAGA 661

Db    1208 GAACGGTCCTGCAGTTAAAAGAACTTTTAAATGGCCTAGGATTCTCCACCCCAGATGAGA 1267

Qy     662 AATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGGA 721

Db    1268 AGTTCCAAAAAGACCCCCCATACAAATGGATGGGCTATGAACTATGGCCAACCAAATGGA 1327

Qy     722 AGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAGT 781
```

```
Db   1328 AGCTGCAAAAAATACAATTGCCCCAAAAAGAAGTATGGACAGTCAATGACATCCAAAAGC 1387

Qy    782 TAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTCT 841

Db   1388 TAGTAGGTGTCCTAAATTGGGCAGCACAAATCTACCCAGGGATAAAGACCAAACACTTAT 1447

Qy    842 GTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCAG 901

Db   1448 GTAGGCTAATTAGAGGAAAAATGACACTCACGGAAGAAGTACAGTGGACAGAACTAGCAG 1507

Qy    902 AAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTACC 961

Db   1508 AGGCAGAACTAGAAGAGAACAAAATTATCTTGAGCCAGGAACAGGAGGGATGCTATTACC 1567

Qy    962 AAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTATA 1021

Db   1568 AAGAAGAAAAGGAATTAGAAGCAACAGTCCAAAAGGATCAAGACAATCAGTGGACATATA 1627

Qy   1022 AAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACAC 1081

Db   1628 AAATACACCAAGGAGAGAAAATCCTAAAAGTAGGAAAATATGCAAAGATAAAAAATACCC 1687

Qy   1082 ATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATAG 1141

Db   1688 ATACCAATGGGGTCAGATTGTTAGCACATGTAGTTCAAAAAATAGGAAAAGAAGCACTAG 1747

Qy   1142 TGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAGT 1201

Db   1748 TCATTTGGGGACGAATACCAAAATTTCACCTACCAGTAGAAAGAGAAACCTGGGAGCAGT 1807

Qy   1202 GGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCAC 1261

Db   1808 GGTGGGATAACTATTGGCAAGTGACATGGATCCCAGACTGGGACTTCGTATCTACTCCAC 1867

Qy   1262 CGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTATT 1321

Db   1868 CACTGGTCAGGTTAGCATTTAACCTAGTAAAAGATCCCATACCAGGTGAAGAGACCTTCT 1927

Qy   1322 ATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGATA 1381

Db   1928 ACACAGATGGATCCTGTAATAGGCAATCAAAAGAGGGAAAAGCAGGATATATAACAGATA 1987

Qy   1382 GGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAAG 1441

Db   1988 GAGGGAGAGACAAGGTAAGGATATTGGAGCAAACTACCAATCAGCAAGCAGAATTAGAAG 2047

Qy   1442 CATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAAT 1501

Db   2048 CCTTCGCAATGGCATTAACAGACTCAGGTCCAAAAGCCAATATTATAGTAGACTCACAGT 2107

Qy   1502 ATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAAA 1561

Db   2108 ATGTAATGGGAATAGTAGCGGGCCAGCCAACAGAATCAGAGAGTAAACTAGTAAACCAAA 2167

Qy   1562 TAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAAG 1621

Db   2168 TCATAGAAGAAATGATAAAAAAGGAAACACTCTATGTTGCATGGGTCCCAGCCCACAAAG 2227

Qy   1622 GTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTCT 1681

Db   2228 GCATAGGAGGAAATCAGGAAGTAGATCATTTAGTAAGTCAGGGCATTAGACAAGTATTAT 2287

Qy   1682 TCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAAG 1741

Db   2288 TCCTAGAAAAAATAGAGCCCGCTCAGGAAGAACATGAGAAATATCATAGCAATGTAAAAG 2347

Qy   1742 AATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGTG 1801

Db   2348 AATTATCCCATAAATTTGGACTGCCCAAACTAGTGGCAAGACAAATAGTAAACACATGTG 2407

Qy   1802 ATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACTT 1861

Db   2408 CCCAATGTCAACAGAAAGGGGAAGCTATACATGGGCAAGTAGATGCAGAACTGGGCACTT 2467
```

```
Qy   1862 GGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCTA 1921

Db   2468 GGCAAATGGACTGCACACACTTAGAGGGAAAAATCATTATAGTAGCAGTACATGTTGCAA 2527

Qy   1922 GTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTTC 1981

Db   2528 GCGGGTTTATAGAAGCAGAAGTTATCCCACAGGAAACGGGAAGGCAAACAGCACTCTTCC 2587

Qy   1982 TGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAACT 2041

Db   2588 TATTAAAACTGGCCAGTAGGTGGCCAATAACACACCTGCACACAGATAATGGTGCCAACT 2647

Qy   2042 TTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGGG 2101

Db   2648 TCACCTCACAGGAAGTAAAGATGGTAGCGTGGTGGACAGGTATAGAACAATCCTTTGGAG 2707

Qy   2102 TACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAATC 2161

Db   2708 TACCTTACAATCCACAAAGCCAAGGAGTAGTAGAAGCAATGAATCACCACTTAAAAAACC 2767

Qy   2162 AAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTTC 2221

Db   2768 AGATAAGCAGAATTAGAGAGCAGGCAAATACAATGGAAACAATAGTATTAATGGCAGTTC 2827

Qy   2222 ATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTAA 2281

Db   2828 ATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACCCCAGCAGAAAGACTAA 2887

Qy   2282 TTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTTA 2341

Db   2888 TCAATATGATCACCACAGAACAAGAAATACAATTCCTCCACGCAAAAAATTCAAAATTAA 2947

Qy   2342 AAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAGC 2401

Db   2948 AAAATTTCCGGGTCTATTTCAGAGAAGGCAGAGATCAGCTGTGGAAAGGACCTGGGGAAC 3007

Qy   2402 TATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTAC 2461

Db   3008 TACTGTGGAAGGGAGATGGAGCAGTCATAGTCAAGGTAGGGACAGACATAAAAATAGTGC 3067

Qy   2462 CCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGTT 2521

Db   3068 CAAGAAGGAAAGCTAAGATCATCAGAGACTATGGAGGAAGGCGAGAGGTGGATAGTAGTT 3127

Qy   2522 CCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAG 2562

Db   3128 CCCACTTGGAGGGTACCAGGGAGGATGGAGAAGTGGCATAG 3168
```

```
RESULT 5
US-09-256-490-14
; Sequence 14, Application US/09256490
; Patent No. 6235881
;  GENERAL INFORMATION:
;    APPLICANT:  Kraus, Guenter
;    APPLICANT:  Wong-Staal, Flossie
;    APPLICANT:  Talbott, Randy
;    APPLICANT:  Poeschla, Eric
;    TITLE OF INVENTION:  Isolation of No. 6235881el HIV-2 Proviruses
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew LLP
;      STREET:  Two Embarcadero Center, Eighth Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/256,490
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/659,251
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Garrett-Wackowski, Eugenia
;      REGISTRATION NUMBER:  37,330
;      REFERENCE/DOCKET NUMBER:  02307E-056410US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 576-0200
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3168 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    FEATURE:
;      NAME/KEY:  -
;      LOCATION:  1..3168
;      OTHER INFORMATION:  /label= pol
;      OTHER INFORMATION:  /note= "HIV-2KR subsequence encoding
;      OTHER INFORMATION:  pol gene"
US-09-256-490-14

  Query Match             68.3%;  Score 1751.4;  DB 4;  Length 3168;
  Best Local Similarity   80.2%;  Pred. No. 0;
  Matches 2055;  Conservative    0;  Mismatches  506;  Indels    0;  Gaps    0;

Qy       2 TTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCAA 61
Db     608 TACCAGTCGCCAAGGTAGACCCGATAAAAGTAATACTGAAACCAGGAAAAGATGGACCAA 667

Qy      62 AATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAAA 121
Db     668 AAGTAAGACAATGGCCTCTAACAAAAGAAAAGATAGAGGCACTAAAAGAAATCTGTGAAA 727

Qy     122 AGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCCA 181
Db     728 AAATGGAAAGAGAAGGCCAGCTAGAGGAAGCTCCCCCAACTAATCCTTATAATACCCCCA 787

Qy     182 CATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAAC 241
Db     788 CATTTGCAATTAAGAAAAAGGACAAAAACAAATGGAGAATGCTAATAGATTTTAGAGAAC 847

Qy     242 TAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGAC 301
Db     848 TAAATAAGGTAACTCAAGAGTTCACAGAAATTCAGTTAGGAATTCCACACCCAGCAGGAT 907

Qy     302 TAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCTC 361
Db     908 TAGCCAAGAAAAGAAGAATTACTGTACTAGATATAGGGGATGCCTACTTTTCCATACCAC 967

Qy     362 TAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAGC 421
Db     968 TACATGAGGACTTTAGACAATATACTGCATTTACTCTACCAACAGTGAACAATGCAGAAC 1027

Qy     422 CAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATCT 481
Db    1028 CAGGAAAGAGATATATATATAAAGTCCTACCACAGGGATGGAAAGGATCGCCAGCAATTT 1087

Qy     482 TCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACCT 541
Db    1088 TTCAACACACAATGAGGCAGGTCTTAGAGCCATTCAGAAAAGCAAACCCAGACGTCATTC 1147

Qy     542 TAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGACA 601
Db    1148 TCGTCCAATATATGGATGATATCTTAATAGCTAGCGACAGGACAGACTTAGAGCATGACA 1207

Qy     602 GGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAGA 661
Db    1208 GAACGGTCCTGCAGTTAAAAGAACTTTTAAATGGCCTAGGATTCTCCACCCCAGATGAGA 1267

Qy     662 AATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGGA 721
Db    1268 AGTTCCAAAAAGACCCCCCATACAAATGGATGGGCTATGAACTATGGCCAACCAAATGGA 1327

Qy     722 AGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAGT 781
Db    1328 AGCTGCAAAAAATACAATTGCCCCAAAAAGAAGTATGGACAGTCAATGACATCCAAAAGC 1387

Qy     782 TAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTCT 841
Db    1388 TAGTAGGTGTCCTAAATTGGGCAGCACAAATCTACCCAGGGATAAAGACCAAACACTTAT 1447

Qy     842 GTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCAG 901
Db    1448 GTAGGCTAATTAGAGGAAAAATGACACTCACGGAAGAAGTACAGTGGACAGAACTAGCAG 1507

Qy     902 AAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTACC 961
Db    1508 AGGCAGAACTAGAAGAGAACAAAATTATCTTGAGCCAGGAACAGGAGGGATGCTATTACC 1567

Qy     962 AAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTATA 1021
Db    1568 AAGAAGAAAAGGAATTAGAAGCAACAGTCCAAAAGGATCAAGACAATCAGTGGACATATA 1627

Qy    1022 AAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACAC 1081
Db    1628 AAATACACCAAGGAGAGAAAATCCTAAAAGTAGGAAAATATGCAAAGATAAAAAATACCC 1687

Qy    1082 ATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATAG 1141
Db    1688 ATACCAATGGGGTCAGATTGTTAGCACATGTAGTTCAAAAAATAGGAAAAGAAGCACTAG 1747

Qy    1142 TGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAGT 1201
Db    1748 TCATTTGGGGACGAATACCAAAATTTCACCTACCAGTAGAAAGAGAAACCTGGGAGCAGT 1807

Qy    1202 GGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCAC 1261
Db    1808 GGTGGGATAACTATTGGCAAGTGACATGGATCCCAGACTGGGACTTCGTATCTACTCCAC 1867

Qy    1262 CGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTATT 1321
Db    1868 CACTGGTCAGGTTAGCATTTAACCTAGTAAAAGATCCCATACCAGGTGAAGAGACCTTCT 1927

Qy    1322 ATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGATA 1381
Db    1928 ACACAGATGGATCCTGTAATAGGCAATCAAAAGAGGGAAAAGCAGGATATATAACAGATA 1987

Qy    1382 GGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAAG 1441
Db    1988 GAGGGAGAGACAAGGTAAGGATATTGGAGCAAACTACCAATCAGCAAGCAGAATTAGAAG 2047

Qy    1442 CATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAAT 1501
Db    2048 CCTTCGCAATGGCATTAACAGACTCAGGTCCAAAAGCCAATATTATAGTAGACTCACAGT 2107

Qy    1502 ATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAAA 1561
Db    2108 ATGTAATGGGAATAGTAGCGGGCCAGCCAACAGAATCAGAGAGTAAACTAGTAAACCAAA 2167

Qy    1562 TAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAAG 1621
Db    2168 TCATAGAAGAAATGATAAAAAAGGAAACACTCTATGTTGCATGGGTCCCAGCCCACAAAG 2227

Qy    1622 GTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTCT 1681
Db    2228 GCATAGGAGGAAATCAGGAAGTAGATCATTTAGTAAGTCAGGGCATTAGACAAGTATTAT 2287

Qy    1682 TCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAAG 1741
```

```
Db   2288 TCCTAGAAAAAATAGAGCCCGCTCAGGAAGAACATGAGAAATATCATAGCAATGTAAAAG 2347

Qy   1742 AATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGTG 1801

Db   2348 AATTATCCCATAAATTTGGACTGCCCAAACTAGTGGCAAGACAAATAGTAAACACATGTG 2407

Qy   1802 ATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACTT 1861

Db   2408 CCCAATGTCAACAGAAAGGGGAAGCTATACATGGGCAAGTAGATGCAGAACTGGGCACTT 2467

Qy   1862 GGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCTA 1921

Db   2468 GGCAAATGGACTGCACACACTTAGAGGGAAAAATCATTATAGTAGCAGTACATGTTGCAA 2527

Qy   1922 GTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTTC 1981

Db   2528 GCGGGTTTATAGAAGCAGAAGTTATCCCACAGGAAACGGGAAGGCAAACAGCACTCTTCC 2587

Qy   1982 TGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAACT 2041

Db   2588 TATTAAAACTGGCCAGTAGGTGGCCAATAACACACCTGCACACAGATAATGGTGCCAACT 2647

Qy   2042 TTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGGG 2101

Db   2648 TCACCTCACAGGAAGTAAAGATGGTAGCGTGGTGGACAGGTATAGAACAATCCTTTGGAG 2707

Qy   2102 TACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAATC 2161

Db   2708 TACCTTACAATCCACAAAGCCAAGGAGTAGTAGAAGCAATGAATCACCACTTAAAAAACC 2767

Qy   2162 AAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTTC 2221

Db   2768 AGATAAGCAGAATTAGAGAGCAGGCAAATACAATGGAAACAATAGTATTAATGGCAGTTC 2827

Qy   2222 ATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTAA 2281

Db   2828 ATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACCCCAGCAGAAAGACTAA 2887

Qy   2282 TTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTTA 2341

Db   2888 TCAATATGATCACCACAGAACAAGAAATACAATTCCTCCACGCAAAAAATTCAAAATTAA 2947

Qy   2342 AAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAGC 2401

Db   2948 AAAATTTCCGGGTCTATTTCAGAGAAGGCAGAGATCAGCTGTGGAAAGGACCTGGGGAAC 3007

Qy   2402 TATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTAC 2461

Db   3008 TACTGTGGAAGGGAGATGGAGCAGTCATAGTCAAGGTAGGGACAGACATAAAAATAGTGC 3067

Qy   2462 CCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGTT 2521

Db   3068 CAAGAAGGAAAGCTAAGATCATCAGAGACTATGGAGGAAGGCGAGAGGTGGATAGTAGTT 3127

Qy   2522 CCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAG 2562

Db   3128 CCCACTTGGAGGGTACCAGGGAGGATGGAGAAGTGGCATAG 3168


RESULT 6
PCT-US96-11445-14
; Sequence 14, Application PC/TUS9611445
;  GENERAL INFORMATION:
;    APPLICANT:  The Regents of the University of California
;    TITLE OF INVENTION:  Isolation of Novel HIV-2 Proviruses
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Robbins, Berliner & Carson
;      STREET:  201 N. Figueroa Street, 5th Floor
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  90012-2628
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30.
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/11445
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berliner, Robert
;      REGISTRATION NUMBER:  20,121
;      REFERENCE/DOCKET NUMBER:  5555-399C1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 977-1001
;      TELEFAX:  (213) 977-1003
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3168 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    FEATURE:
;      NAME/KEY:  -
;      LOCATION:  1..3168
;      OTHER INFORMATION:  /label= pol
;      OTHER INFORMATION:  /note= "HIV-2KR subsequence encoding
;      OTHER INFORMATION:  pol gene"
PCT-US96-11445-14

  Query Match             68.3%;  Score 1751.4;  DB 5;  Length 3168;
  Best Local Similarity   80.2%;  Pred. No. 0;
  Matches 2055;  Conservative    0;  Mismatches  506;  Indels    0;  Gaps    0;

Qy       2 TTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCAA 61

Db     608 TACCAGTCGCCAAGGTAGACCCGATAAAAGTAATACTGAAACCAGGAAAAGATGGACCAA 667

Qy      62 AATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAAA 121

Db     668 AAGTAAGACAATGGCCTCTAACAAAAGAAAAGATAGAGGCACTAAAAGAAATCTGTGAAA 727

Qy     122 AGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCCA 181

Db     728 AAATGGAAAGAGAAGGCCAGCTAGAGGAAGCTCCCCCAACTAATCCTTATAATACCCCCA 787

Qy     182 CATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAAC 241

Db     788 CATTTGCAATTAAGAAAAAGGACAAAAACAAATGGAGAATGCTAATAGATTTTAGAGAAC 847

Qy     242 TAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGAC 301

Db     848 TAAATAAGGTAACTCAAGAGTTCACAGAAATTCAGTTAGGAATTCCACACCCAGCAGGAT 907

Qy     302 TAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCTC 361

Db     908 TAGCCAAGAAAAGAAGAATTACTGTACTAGATATAGGGGATGCCTACTTTTCCATACCAC 967

Qy     362 TAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAGC 421

Db     968 TACATGAGGACTTTAGACAATATACTGCATTTACTCTACCAACAGTGAACAATGCAGAAC 1027

Qy     422 CAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATCT 481

Db    1028 CAGGAAAGAGATATATATATAAAGTCCTACCACAGGGATGGAAAGGATCGCCAGCAATTT 1087

Qy     482 TCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACCT 541

Db    1088 TTCAACACACAATGAGGCAGGTCTTAGAGCCATTCAGAAAAGCAAACCCAGACGTCATTC 1147

Qy     542 TAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGACA 601
```

```
Db   1148 TCGTCCAATATATGGATGATATCTTAATAGCTAGCGACAGGACAGACTTAGAGCATGACA 1207

Qy    602 GGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAGA 661

Db   1208 GAACGGTCCTGCAGTTAAAAGAACTTTTAAATGGCCTAGGATTCTCCACCCCAGATGAGA 1267

Qy    662 AATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGGA 721

Db   1268 AGTTCCAAAAAGACCCCCCATACAAATGGATGGGCTATGAACTATGGCCAACCAAATGGA 1327

Qy    722 AGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAGT 781

Db   1328 AGCTGCAAAAAATACAATTGCCCCAAAAAGAAGTATGGACAGTCAATGACATCCAAAAGC 1387

Qy    782 TAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTCT 841

Db   1388 TAGTAGGTGTCCTAAATTGGGCAGCACAAATCTACCCAGGGATAAAGACCAAACACTTAT 1447

Qy    842 GTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCAG 901

Db   1448 GTAGGCTAATTAGAGGAAAAATGACACTCACGGAAGAAGTACAGTGGACAGAACTAGCAG 1507

Qy    902 AAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTACC 961

Db   1508 AGGCAGAACTAGAAGAGAACAAAATTATCTTGAGCCAGGAACAGGAGGGATGCTATTACC 1567

Qy    962 AAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTATA 1021

Db   1568 AAGAAGAAAAGGAATTAGAAGCAACAGTCCAAAAGGATCAAGACAATCAGTGGACATATA 1627

Qy   1022 AAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACAC 1081

Db   1628 AAATACACCAAGGAGAGAAAATCCTAAAAGTAGGAAAATATGCAAAGATAAAAAATACCC 1687

Qy   1082 ATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATAG 1141

Db   1688 ATACCAATGGGGTCAGATTGTTAGCACATGTAGTTCAAAAAATAGGAAAAGAAGCACTAG 1747

Qy   1142 TGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAGT 1201

Db   1748 TCATTTGGGGACGAATACCAAAATTTCACCTACCAGTAGAAAGAGAAACCTGGGAGCAGT 1807

Qy   1202 GGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCAC 1261

Db   1808 GGTGGGATAACTATTGGCAAGTGACATGGATCCCAGACTGGGACTTCGTATCTACTCCAC 1867

Qy   1262 CGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTATT 1321

Db   1868 CACTGGTCAGGTTAGCATTTAACCTAGTAAAAGATCCCATACCAGGTGAAGAGACCTTCT 1927

Qy   1322 ATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGATA 1381

Db   1928 ACACAGATGGATCCTGTAATAGGCAATCAAAAGAGGGAAAAGCAGGATATATAACAGATA 1987

Qy   1382 GGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAAG 1441

Db   1988 GAGGGAGAGACAAGGTAAGGATATTGGAGCAAACTACCAATCAGCAAGCAGAATTAGAAG 2047

Qy   1442 CATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAAT 1501

Db   2048 CCTTCGCAATGGCATTAACAGACTCAGGTCCAAAAGCCAATATTATAGTAGACTCACAGT 2107

Qy   1502 ATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAAA 1561

Db   2108 ATGTAATGGGAATAGTAGCGGGCCAGCCAACAGAATCAGAGAGTAAACTAGTAAACCAAA 2167

Qy   1562 TAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAAG 1621

Db   2168 TCATAGAAGAAATGATAAAAAAGGAAACACTCTATGTTGCATGGGTCCCAGCCCACAAAG 2227

Qy   1622 GTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTCT 1681

Db   2228 GCATAGGAGGAAATCAGGAAGTAGATCATTTAGTAAGTCAGGGCATTAGACAAGTATTAT 2287

Qy   1682 TCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAAG 1741

Db   2288 TCCTAGAAAAAATAGAGCCCGCTCAGGAAGAACATGAGAAATATCATAGCAATGTAAAAG 2347

Qy   1742 AATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGTG 1801

Db   2348 AATTATCCCATAAATTTGGACTGCCCAAACTAGTGGCAAGACAAATAGTAAACACATGTG 2407

Qy   1802 ATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACTT 1861

Db   2408 CCCAATGTCAACAGAAAGGGGAAGCTATACATGGGCAAGTAGATGCAGAACTGGGCACTT 2467

Qy   1862 GGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCTA 1921

Db   2468 GGCAAATGGACTGCACACACTTAGAGGGAAAAATCATTATAGTAGCAGTACATGTTGCAA 2527

Qy   1922 GTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTTC 1981

Db   2528 GCGGGTTTATAGAAGCAGAAGTTATCCCACAGGAAACGGGAAGGCAAACAGCACTCTTCC 2587

Qy   1982 TGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAACT 2041

Db   2588 TATTAAAACTGGCCAGTAGGTGGCCAATAACACACCTGCACACAGATAATGGTGCCAACT 2647

Qy   2042 TTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGGG 2101

Db   2648 TCACCTCACAGGAAGTAAAGATGGTAGCGTGGTGGACAGGTATAGAACAATCCTTTGGAG 2707

Qy   2102 TACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAATC 2161

Db   2708 TACCTTACAATCCACAAAGCCAAGGAGTAGTAGAAGCAATGAATCACCACTTAAAAAACC 2767

Qy   2162 AAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTTC 2221

Db   2768 AGATAAGCAGAATTAGAGAGCAGGCAAATACAATGGAAACAATAGTATTAATGGCAGTTC 2827

Qy   2222 ATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTAA 2281

Db   2828 ATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACCCCAGCAGAAAGACTAA 2887

Qy   2282 TTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTTA 2341

Db   2888 TCAATATGATCACCACAGAACAAGAAATACAATTCCTCCACGCAAAAAATTCAAAATTAA 2947

Qy   2342 AAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAGC 2401

Db   2948 AAAATTTCCGGGTCTATTTCAGAGAAGGCAGAGATCAGCTGTGGAAAGGACCTGGGGAAC 3007

Qy   2402 TATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTAC 2461

Db   3008 TACTGTGGAAGGGAGATGGAGCAGTCATAGTCAAGGTAGGGACAGACATAAAAATAGTGC 3067

Qy   2462 CCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGTT 2521

Db   3068 CAAGAAGGAAAGCTAAGATCATCAGAGACTATGGAGGAAGGCGAGAGGTGGATAGTAGTT 3127

Qy   2522 CCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAG 2562

Db   3128 CCCACTTGGAGGGTACCAGGGAGGATGGAGAAGTGGCATAG 3168


RESULT 7
5223423-1
;Patent No. 5223423
;     APPLICANT: FRANCHINI, GENOVEFFA;WONG-STAAL, FLOSSIE;
;GALLO, ROBERT
;     TITLE OF INVENTION: CHARACTERIZATION OF REPLICATION COMPETENT
;HUMAN IMMUNODEFICIENCY TYPE 2 PROVIRAL CLONE HIV-2 SBL/ISY
;     NUMBER OF SEQUENCES: 4
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/331,212
```

```
;      FILING DATE: 03-31-1989
;SEQ ID NO:1:
;      LENGTH: 9633
5223423-1

  Query Match             67.8%;  Score 1738.4;  DB 6;  Length 9633;
  Best Local Similarity   79.9%;  Pred. No. 0;
  Matches 2048;  Conservative    0;  Mismatches  516;  Indels    0;  Gaps    0;

Qy       1 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 60
Db    2373 TTACCAGTCGCCAAAATAGAACCAGTAAAAGTAACATTAAAGCCAGGAAAAGATGGGCCA 2432

Qy      61 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 120
Db    2433 AAACAAAGACAATGGCCCTTAACAAGAGAAAAAATAGAAGCACTAAGAGAAATCTGTGAA 2492

Qy     121 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 180
Db    2493 AAAATGGAAAGAGAAGGTCAGCTAGAAGAAGCGCCTCCAACTAATCCCTATAATACCCCT 2552

Qy     181 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 240
Db    2553 ACATTTGCAATTAAGAAAAAGGACAAAAACAAATGGAGGATGCTGATAGATTTTAGAGAA 2612

Qy     241 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 300
Db    2613 CTAAACAAGGTAACTCAAGATTTCACAGAGGTTCAGTTAGGAATTCCACACCCAGCAGGA 2672

Qy     301 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 360
Db    2673 TTAGCCAAGAAAAGAAGAATTACTGTGTTAGATGTAGGAGATGCCTACTTTTCCATACCC 2732

Qy     361 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 420
Db    2733 CTATATGAGGATTTTAGACAGTATACTGCATTTACTCTGCCATCAGTAAACAATGCAGAA 2792

Qy     421 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 480
Db    2793 CCAGGAAAAAGATATATATACAAAGTCTTACCACAGGGATGGAAGGGGTCACCAGCAATT 2852

Qy     481 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 540
Db    2853 TTTCAATACACAATGAGGCAAGTCTTAGAACCATTCAGAAAAGCAAACCCAGATGTCATT 2912

Qy     541 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 600
Db    2913 ATCGTTCAGTACATGGATGATATCTTAATAGCTAGTGACAGGACAGATTTGGAACATGAC 2972

Qy     601 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 660
Db    2973 AAAGTAGTCCTACAGCTAAAGGAACTTCTAAATGGCCTGGGATTTTCCACCCCAGACGAA 3032

Qy     661 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 720
Db    3033 AAGTTCCAAAAGGACCCTCCATACCAATGGATGGGCTATGAACTGTGGCCAACCAAATGG 3092

Qy     721 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 780
Db    3093 AAATTGCAAAAAATACAATTGCCCCAAAAGGAAGTATGGACAGTTAATGACATCCAGAAA 3152

Qy     781 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 840
Db    3153 CTAGTGGGTGTCCTAAACTGGGCGGCACAAATCTACCCAGGAATAAAGACCAAACACTTA 3212

Qy     841 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 900
Db    3213 TGTAAGCTAATTAGAGGAAAGATGACACCCACGGAAGAAGTACAGTGGACAGAATTAGCA 3272

Qy     901 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 960
Db    3273 GAAGCAGAGCTGGAGGAAAACAAAATTATCTTAAGCCAGGAACAGGAGGGACACTATTAC 3332

Qy     961 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 1020
Db    3333 CAAGAGGAAAAAGAGTTAGAAGCAACAGTTCAAAAGGATCAAGACAATCAGTGGACATAT 3392

Qy    1021 AAAATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAATACA 1080
Db    3393 AAAGTACACCAGGGAGAAAAAATTCTAAAAGTAGGAAAATATGCAAAGATAAAAAATACC 3452

Qy    1081 CATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATA 1140
Db    3453 CATACCAACGGGGTCAGGTTGTTGGCACAGGTAGTTCAGAAAATAGGAAAAGAAGCACTA 3512

Qy    1141 GTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAG 1200
Db    3513 GTCATTTGGGGACGAATACCAAAATTTCACCTACCAGTAGAAAGAGAGACCTGGGAACAG 3572

Qy    1201 TGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCA 1260
Db    3573 TGGTGGGATAACTATTGGCAAGTGACATGGATCCCAGACTGGGACTTCGTATCCACCCCA 3632

Qy    1261 CCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTAT 1320
Db    3633 CCGTTGGTCAGGTTAGCATTTAACCTGGTAAAAGATCCTATACCAGGCGCAGAGACCTTC 3692

Qy    1321 TATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGAT 1380
Db    3693 TACACGGATGGATCTTGCAATAGGCAATCAAAAGAGGGAAAAGCAGGATATATAACAGAT 3752

Qy    1381 AGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAA 1440
Db    3753 AGAGGAAAAGACAAAGTAAGGATATTAGAGCAAACTACCAACCAACAAGCAGAATTAGAA 3812

Qy    1441 GCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAA 1500
Db    3813 GCCTTTGCAATGGCAGTAACAGACTCAGGTCCAAAAGTCAATATTGTAGTAGACTCACAG 3872

Qy    1501 TATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAA 1560
Db    3873 TATGTAATGGGAATAGTAACAGGCCAACCGGCTGAATCAGAGAGTAGAATAGTAAATAAA 3932

Qy    1561 ATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAA 1620
Db    3933 ATTATAGAAGAGATGATAAAAAAGGAAGCAATCTATGTTGCATGGGTCCCGGCCCACAAA 3992

Qy    1621 GGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTC 1680
Db    3993 GGCATAGGAGGAAATCAAGAAATTGACCACTTAGTAAGTCAGGGCATCAGACAAGTATTA 4052

Qy    1681 TTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAA 1740
Db    4053 TTCCTAGAGAGAATAGAGCCCGCTCAGGAAGAACATGGAAAATATCATAGCAATGTAAAA 4112

Qy    1741 GAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGT 1800
Db    4113 GAACTAGCCCATAAGTTTGGATTACCCAACCTGGTGGCAAGACAAATAGTAAACACATGT 4172

Qy    1801 GATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACT 1860
Db    4173 GCCCAGTGCCAACAAAAAGGGGAAGCTATACATGGGCAAGTAAATGCAGAACTAGGCACC 4232

Qy    1861 TGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCT 1920
Db    4233 TGGCAAATGGACTGCACACACTTAGAAGGAAAAATCATTATAGTAGCAGTACATGTTGCA 4292

Qy    1921 AGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTT 1980
Db    4293 AGTGGATTTATAGAAGCAGAAGTCATCCCACAGGAATCAGGAAGGCAAACAGCACTCTTC 4352

Qy    1981 CTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAAC 2040
Db    4353 CTATTAAAACTGGCCAGTAGGTGGCCAATAACACACTTGCACACAGATAATGGTGCCAAC 4412

Qy    2041 TTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGG 2100
```

```
Db    4413 TTCACTTCACAGGAGGTAAAGATGGTAGCATGGTGGGTAGGCATAGAACAATCCTTTGGA 4472

Qy    2101 GTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAAT 2160

Db    4473 GTACCTTACAATCCACAAAGCCAGGGAGTAGTAGAAGCAATGAATCACCACCTGAAAAAT 4532

Qy    2161 CAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTT 2220

Db    4533 CAGATAGAAAGAATTAGAGAGCAGGCAAATACAATGGAAACAATAGTACTAATGGCAGTT 4592

Qy    2221 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTA 2280

Db    4593 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACCCCAGTAGAAAGACTA 4652

Qy    2281 ATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTT 2340

Db    4653 GTCAATATGATCACCACAGAACAAGAAATACAATTCCTCCAAGCAAAAAATTCAAAATTA 4712

Qy    2341 AAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAG 2400

Db    4713 AAAAATTTTCGGGTCTATTTCAGAGAAGGCAGAAATCAACTGTGGCAAGGACCTGGGGAG 4772

Qy    2401 CTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTA 2460

Db    4773 CTACTGTGGAAAGGGGACGGAGCAGTCATAGTCAAGGTAGGGACAGATATAAAAGTAATA 4832

Qy    2461 CCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGATAGCAGT 2520

Db    4833 CCAAGAAGAAAGGCCAAGATCATCAGAGACTATGGACCAAGGCAAGAGATGGATAGCGGT 4892

Qy    2521 TCCCACATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCC 2564

Db    4893 TCCCACCTGGAGGGTGCCAGGGAGGATGGAGAAATGGCATAGCC 4936


RESULT 8
US-08-876-546A-15
; Sequence 15, Application US/08876546A
; Patent No. 5962665
;  GENERAL INFORMATION:
;    APPLICANT:  P. Kroeger
;    APPLICANT:  K. Abravaya
;    APPLICANT:  J. Gorzowski
;    APPLICANT:  R. Hoenle
;    APPLICANT:  C. Esping
;    APPLICANT:  J. Moore
;    TITLE OF INVENTION:  NUCLEIC ACID SEQUENCES FOR DETECTING
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Road
;      CITY:  Abbott Park
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC Compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/876,546A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Paul D. Yasger
;      REGISTRATION NUMBER:  37,477
;      REFERENCE/DOCKET NUMBER:  6127.US.01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  847/938-3508
;      TELEFAX:  847/938-2623
;      TELEX:
```

```
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2689 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic RNA (HIV-2)
US-08-876-546A-15

  Query Match            61.9%;  Score 1587;  DB 2;  Length 2689;
  Best Local Similarity  79.6%;  Pred. No. 0;
  Matches 1900;  Conservative    0;  Mismatches  485;  Indels    2;  Gaps    2;

Qy      2 TTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCAA 61

Db    304 TGCCAGTTGCCAAGATAGAACCAATAAAAATAATGCTAAAGCCAGGGAAAGATGGACCAA 363

Qy     62 AATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAAA 121

Db    364 GACTAAAACAATGGCCTTTAACAAAAGAAAAAATAGAAGCACTAAAAGAAATCTGTGAAA 423

Qy    122 AGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCCA 181

Db    424 AAATGGAAAAAGAAGGCCAGCTAGAAGAAGCACCTCCAACTAATCCTTATAATACCCCCA 483

Qy    182 CATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAAC 241

Db    484 CATTTGCAATCAAGAAAAAGGACAAAAACAAATGGAGAATGCTAATAGACTTTAGGGAGC 543

Qy    242 TAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGAC 301

Db    544 TAAACAAGGTAACTCAGGATTTCACAGAAATTCAGTTAGGAATTCCACACCCAGCAGGAT 603

Qy    302 TAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCTC 361

Db    604 TGGCCAAAAAGAGGAGAATTACTGTACTAGATGTAGGGGATGCTTACTTTTCCATACCAC 663

Qy    362 TAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAGC 421

Db    664 TACATGAGGACTTTAGACAGTATACTGCATTCACTCTACCATCAGTAAACAATGCAGAAC 723

Qy    422 CAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATCT 481

Db    724 CAGGAAAAAGATATATATACAAAGTCCTACCACAGGGGTGGAAGGGGTCACCAGCAATTT 783

Qy    482 TCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACCT 541

Db    784 TTCAATACACAATGAGGCAGATCTTAGAACCATTCAGAAAAGCAAACGAGGATGTCATTA 843

Qy    542 TAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGACA 601

Db    844 TCATTCAGTACATGGATGATATCTTAATAGCCAGCGACAGGACAGACTTAGAACATGACA 903

Qy    602 GGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAGA 661

Db    904 AAGTGGTCCTGCAGCTAAAAGAACTTCTAAATGGACTAGGATTTTCCACCCCAGATGAGA 963

Qy    662 AATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGGA 721

Db    964 AATTCCAGAAAGACCCTCCATATCGCTGGATGGGCTATGAATTATGGCCAACTAAATGGA 1023

Qy    722 AGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAGT 781

Db   1024 AGTTGCAAAAAATACAGCTGCCCCAAAAAGAAGTATGGACAGTCAATGACATCCAAAAGC 1083

Qy    782 TAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTCT 841

Db   1084 TAGTGGGTGTCCTAAATTGGGCAGCACAAATCTACCCAGGGATAAAGACCAAACACCTAT 1143

Qy    842 GTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCAG 901

Db   1144 GTAGGCTAATTAGAGGAAAAATGACACTCACAGAAGAGGTACAGTGGACAGAATTAGCAG 1203
```

```
Qy    902 AAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTACC 961
          |||||||    ||||||||| | ||| ||  | || ||| ||||||||||||   |||||||
Db   1204 AAGCAGAGCTAGAGGAAAACAGAATTATCTTAAGCCAGAAACAAGAAGGACACTATTACC 1263

Qy    962 AAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTATA 1021
          | ||||  ||    |||||||| || ||    ||   ||| |||||||||||| | ||||
Db   1264 AGGAAGAAAAAAAGTTAGAAGCAACAGTCCAGAAAGATCAAGACAATCAGTGGACATATA 1323

Qy   1022 AAATTCACCAAGAAGACAAAATACTGAAA-GTAGGAAAATTTGCAAAGATAAAGAATACA 1080
          |  | |||||  |  || ||||| || ||| || ||||||| |||||||||||| |||||
Db   1324 AGGTACACCAGGGGGAGAAAATTCTCAAAGGTGGGAAAATATGCAAAGATAAAAAATACC 1383

Qy   1081 CATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATA 1140
          |||||||| || || ||| | |||||||| ||| | || |||||||||| |||||| ||
Db   1384 CATACCAACGG-GTCAGATTGTTAGCACAGGTAGTTCAAAAAATAGGAAAAGAAGCACTA 1442

Qy   1141 GTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAG 1200
          || || ||||||| | | |||||||| ||| ||||||| |||  |||   ||||| |||
Db   1443 GTCATTTGGGGACGGATACCAAAATTTCACCTACCAGTAGAGAGAGATACCTGGGAGCAG 1502

Qy   1201 TGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCA 1260
          ||||||    |||| ||||| ||||| ||||| || || ||||| ||| | || || |||
Db   1503 TGGTGGGATAACTACTGGCAAGTAACATGGATCCCAGACTGGGACTTTGTATCTACCCCA 1562

Qy   1261 CCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTAT 1320
          || || || ||  |||  || || ||  ||  ||| ||   || | ||||||  |
Db   1563 CCACTGGTCAGGCTAGCATTTAACCTGGTAGGAGAGCCTGTACCAGGCGCAGAAACTTTC 1622

Qy   1321 TATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGAT 1380
          || |||||||||| || ||||  |||||||||||||||||||||||||||| ||||||
Db   1623 TACACAGATGGATCCTGCAATAGGCAGTCAAAAGAAGGGAAAGCAGGATATATAACAGAT 1682

Qy   1381 AGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAA 1440
          || || | |||||  |||||||| ||||| || ||||| |||||  |||||||||| |||
Db   1683 AGAGGGAGAGACAGGGTAAAAGTATTAGAGCAAACTACCAATCAGCAAGCAGAATTAGAA 1742

Qy   1441 GCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAA 1500
          || |||   |||||| | |||||||||| ||||| || |||||||||||| ||||||
Db   1743 GCCTTTGCAATGGCACTAACAGACTCAGGTCCAAAAGCTAATATTATAGTAGACTCACAA 1802

Qy   1501 TATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAA 1560
          ||||| ||||| ||| || ||||    || ||||| |||||||  ||  |||| |||||
Db   1803 TATGTAATGGGGATAGTAGCAGGCCAACCAACAGAGTCAGAGAATAGAATAGTAAATCAG 1862

Qy   1561 ATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAA 1620
          || |||||||||||||| || ||   || ||| ||||| || ||||| ||||| ||||||
Db   1863 ATCATAGAAGAAATGATAAAGAAAGAAGCAATCTATGTTGCGTGGGTCCCAGCCCACAAA 1922

Qy   1621 GGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTC 1680
          || |||||||||||||| ||| |||| ||  |||| |||||.|| ||||||||||   |
Db   1923 GGCATAGGAGGAAACCAGGAAGTAGATCATTTAGTAAGTCAGGGCATTAGACAAGTATTA 1982

Qy   1681 TTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAA 1740
          ||| | || || ||||| || || || |||||||||| ||||| ||||| ||| |||||
Db   1983 TTCCTAGAGAAAATAGAACCCGCGCAGGAAGAACATGAAAAATATCATAGCAATATAAAA 2042

Qy   1741 GAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGT 1800
          ||| ||      |||||||| ||||||   ||||||| ||||| |||||| |||| |||
Db   2043 GAACTGTCCCATAAATTTGGAATACCCAAGCTAGTGGCAAGACAAATAGTAAACACATGT 2102

Qy   1801 GATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACT 1860
          |   |    || |||||||| ||||||||||||||| | |||| |||| ||||| |||
Db   2103 GCCCATGTGCAACAGAAAGGGGAGGCTATACATGGGCAAGTAAATGCAGAACTAGGCACT 2162

Qy   1861 TGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCT 1920
          |||||||||| || || ||| |||| |||||| | || ||||| |||||||||| ||
Db   2163 TGGCAAATGGACTGCACACATTTAGAAGGAAAAGTCATTATAGTAGCAGTACATGTTGCA 2222

Qy   1921 AGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTT 1980
          |||||||| ||||| ||||| || || ||||| ||  ||||||| || |||||||| ||
Db   2223 AGTGGATTTATAGAGGCAGAAGTTATCCCACAGGAATCAGGAAGGCAAACAGCACTGTTC 2282

Qy   1981 CTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAAC 2040
```

```
          |  | ||| ||||  | || ||||| || ||||| || |||||||||||||||| |||
Db   2283 TTACTGAAACTGGCCAGTAGGTGGCCAATAACACACCTGCACACAGATAATGGTGCCAAC 2342

Qy   2041 TTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGG 2100
          ||  |||| || ||||| || ||||| |||||||||| ||| ||||| || |||||||||
Db   2343 TTCACTTCACAGGAAGTGAAAATGGTAGCATGGTGGGTAGGTATAGAACAAACCTTTGGA 2402

Qy   2101 GTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAAT 2160
          || || |||||||||| || || |||||||| ||||||||||||| || || ||||||
Db   2403 GTGCCTTACAATCCACAAAGCCAAGGAGTAGTAGAAGCAATGAATCATCATCTAAAAAAT 2462

Qy   2161 CAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTT 2220
          || ||||| ||||| || || || |||||| |||||||||| ||||||||||||||||||
Db   2463 CAGATAGACAGAATTAGAGAGCAGGCAAATACAGTAGAAACAATAGTATTAATGGCAGTT 2522

Qy   2221 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTA 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||| ||
Db   2523 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACCCCAGCAGAAAGAATA 2582

Qy   2281 ATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTT 2340
          || || |||||||| |||||||||| |||||||  |  ||| ||||||| ||||||||
Db   2583 ATCAATATGATCACCACAGAACAAGAAATACAATTCCTCCAAGCAAAAAATTCAAAATTA 2642

Qy   2341 AAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAA 2387
          |||||||||||||||||| |||||||||||||||||| ||||||||
Db   2643 AAAAATTTTCGGGTCTATTTCAGAGAAGGCAGAGATCAGCTGTGGAA 2689
```

```
RESULT 9
US-09-412-252-15
; Sequence 15, Application US/09412252
; Patent No. 6232455
;  GENERAL INFORMATION:
;    APPLICANT:  P. Kroeger
;    APPLICANT:  K. Abravaya
;    APPLICANT:  J. Gorzowski
;    APPLICANT:  R. Hoenle
;    APPLICANT:  C. Esping
;    APPLICANT:  J. Moore
;    TITLE OF INVENTION:  NUCLEIC ACID SEQUENCES FOR DETECTING
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Road
;      CITY:  Abbott Park
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC Compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/412,252
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/876,546
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Paul D. Yasger
;      REGISTRATION NUMBER:  37,477
;      REFERENCE/DOCKET NUMBER:  6127.US.01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  847/938-3508
;      TELEFAX:  847/938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2689 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
```

```
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic RNA (HIV-2)
US-09-412-252-15

  Query Match             61.9%;  Score 1587;  DB 4;  Length 2689;
  Best Local Similarity   79.6%;  Pred. No. 0;
  Matches 1900;  Conservative    0;  Mismatches  485;  Indels    2;  Gaps    2;

Qy       2 TTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCAA 61
Db     304 TGCCAGTTGCCAAGATAGAACCAATAAAAATAATGCTAAAGCCAGGGAAAGATGGACCAA 363

Qy      62 AATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAAA 121
Db     364 GACTAAAACAATGGCCTTTAACAAAAGAAAAAATAGAAGCACTAAAAGAAATCTGTGAAA 423

Qy     122 AGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCCA 181
Db     424 AAATGGAAAAAGAAGGCCAGCTAGAAGAAGCACCTCCAACTAATCCTTATAATACCCCCA 483

Qy     182 CATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAAC 241
Db     484 CATTTGCAATCAAGAAAAAGGACAAAAACAAATGGAGAATGCTAATAGACTTTAGGGAGC 543

Qy     242 TAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGAC 301
Db     544 TAAACAAGGTAACTCAGGATTTCACAGAAATTCAGTTAGGAATTCCACACCCAGCAGGAT 603

Qy     302 TAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCTC 361
Db     604 TGGCCAAAAAGAGGAGAATTACTGTACTAGATGTAGGGGATGCTTACTTTTCCATACCAC 663

Qy     362 TAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAGC 421
Db     664 TACATGAGGACTTTAGACAGTATACTGCATTCACTCTACCATCAGTAAACAATGCAGAAC 723

Qy     422 CAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATCT 481
Db     724 CAGGAAAAAGATATATATACAAAGTCCTACCACAGGGGTGGAAGGGGTCACCAGCAATTT 783

Qy     482 TCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACCT 541
Db     784 TTCAATACACAATGAGGCAGATCTTAGAACCATTCAGAAAAGCAAACGAGGATGTCATTA 843

Qy     542 TAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGACA 601
Db     844 TCATTCAGTACATGGATGATATCTTAATAGCCAGCGACAGGACAGACTTAGAACATGACA 903

Qy     602 GGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAGA 661
Db     904 AAGTGGTCCTGCAGCTAAAAGAACTTCTAAATGGACTAGGATTTTCCACCCCAGATGAGA 963

Qy     662 AATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGGA 721
Db     964 AATTCCAGAAAGACCCTCCATATCGCTGGATGGGCTATGAATTATGGCCAACTAAATGGA 1023

Qy     722 AGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAGT 781
Db    1024 AGTTGCAAAAAATACAGCTGCCCCAAAAAGAAGTATGGACAGTCAATGACATCCAAAAGC 1083

Qy     782 TAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTCT 841
Db    1084 TAGTGGGTGTCCTAAATTGGGCAGCACAAATCTACCCAGGGATAAAGACCAAACACCTAT 1143

Qy     842 GTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCAG 901
Db    1144 GTAGGCTAATTAGAGGAAAAATGACACTCACAGAAGAGGTACAGTGGACAGAATTAGCAG 1203

Qy     902 AAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTACC 961
Db    1204 AAGCAGAGCTAGAGGAAAACAGAATTATCTTAAGCCAGAAACAAGAAGGACACTATTACC 1263

Qy     962 AAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTATA 1021

Db    1264 AGGAAGAAAAAAAGTTAGAAGCAACAGTCCAGAAAGATCAAGACAATCAGTGGACATATA 1323

Qy    1022 AAATTCACCAAGAAGACAAAATACTGAAA-GTAGGAAAATTTGCAAAGATAAAGAATACA 1080
Db    1324 AGGTACACCAGGGGGAGAAAATTCTCAAAGGTGGGAAAATATGCAAAGATAAAAAATACC 1383

Qy    1081 CATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATA 1140
Db    1384 CATACCAACGG-GTCAGATTGTTAGCACAGGTAGTTCAAAAAATAGGAAAAGAAGCACTA 1442

Qy    1141 GTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAACAG 1200
Db    1443 GTCATTTGGGGACGGATACCAAAATTTCACCTACCAGTAGAGAGAGATACCTGGGAGCAG 1502

Qy    1201 TGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACACCA 1260
Db    1503 TGGTGGGATAACTACTGGCAAGTAACATGGATCCCAGACTGGGACTTTGTATCTACCCCA 1562

Qy    1261 CCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTAT 1320
Db    1563 CCACTGGTCAGGCTAGCATTTAACCTGGTAGGAGAGCCTGTACCAGGCGCAGAAACTTTC 1622

Qy    1321 TATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACAGAT 1380
Db    1623 TACACAGATGGATCCTGCAATAGGCAGTCAAAAGAAGGGAAAGCAGGATATATAACAGAT 1682

Qy    1381 AGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTGGAA 1440
Db    1683 AGAGGGAGAGACAGGGTAAAAGTATTAGAGCAAACTACCAATCAGCAAGCAGAATTAGAA 1742

Qy    1441 GCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCACAA 1500
Db    1743 GCCTTTGCAATGGCACTAACAGACTCAGGTCCAAAAGCTAATATTATAGTAGACTCACAA 1802

Qy    1501 TATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAATCAA 1560
Db    1803 TATGTAATGGGGATAGTAGCAGGCCAACCAACAGAGTCAGAGAATAGAATAGTAAATCAG 1862

Qy    1561 ATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACACAAA 1620
Db    1863 ATCATAGAAGAAATGATAAAGAAAGAAGCAATCTATGTTGCGTGGGTCCCAGCCCACAAA 1922

Qy    1621 GGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCTC 1680
Db    1923 GGCATAGGAGGAAACCAGGAAGTAGATCATTTAGTAAGTCAGGGCATTAGACAAGTATTA 1982

Qy    1681 TTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTAAAA 1740
Db    1983 TTCCTAGAGAAAATAGAACCCGCGCAGGAAGAACATGAAAAATATCATAGCAATATAAAA 2042

Qy    1741 GAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACCTGT 1800
Db    2043 GAACTGTCCCATAAATTTGGAATACCCAAGCTAGTGGCAAGACAAATAGTAAACACATGT 2102

Qy    1801 GATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGGACT 1860
Db    2103 GCCCATGTGCAACAGAAAGGGGAGGCTATACATGGGCAAGTAAATGCAGAACTAGGCACT 2162

Qy    1861 TGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTAGCT 1920
Db    2163 TGGCAAATGGACTGCACACATTTAGAAGGAAAAGTCATTATAGTAGCAGTACATGTTGCA 2222

Qy    1921 AGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTATTT 1980
Db    2223 AGTGGATTTATAGAGGCAGAAGTTATCCCACAGGAATCAGGAAGGCAAACAGCACTGTTC 2282

Qy    1981 CTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCTAAC 2040
Db    2283 TTACTGAAACTGGCCAGTAGGTGGCCAATAACACACCTGCACACAGATAATGGTGCCAAC 2342

Qy    2041 TTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTTGGG 2100
```

```
Db    2343 TTCACTTCACAGGAAGTGAAAATGGTAGCATGGTGGGTAGGTATAGAACAAACCTTTGGA 2402

Qy    2101 GTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAAAAT 2160
           || || ||||||||||| || || |||||||| |||||||||||||| || || ||||||
Db    2403 GTGCCTTACAATCCACAAAGCCAAGGAGTAGTAGAAGCAATGAATCATCATCTAAAAAAT 2462

Qy    2161 CAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTT 2220
           || ||||| ||||| || || || |||||| |||||||||| ||||||||||||||||||
Db    2463 CAGATAGACAGAATTAGAGAGCAGGCAAATACAGTAGAAACAATAGTATTAATGGCAGTT 2522

Qy    2221 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTA 2280
           |||||||||||||||||||||||||||||||||||||||||||| ||||||||||| ||
Db    2523 CATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACCCCAGCAGAAAGAATA 2582

Qy    2281 ATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTT 2340
           || || |||||||| |||||||||| |||||||| | ||| ||||||| ||||||||
Db    2583 ATCAATATGATCACCACAGAACAAGAAATACAATTCCTCCAAGCAAAAAATTCAAAATTA 2642

Qy    2341 AAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAA 2387
           |||||||||||||||||| |||||||||||||||| ||||||||
Db    2643 AAAAATTTTCGGGTCTATTTCAGAGAAGGCAGAGATCAGCTGTGGAA 2689


RESULT 10
US-08-022-835-3
; Sequence 3, Application US/08022835
; Patent No. 5420030
;  GENERAL INFORMATION:
;    APPLICANT:  Reitz Jr., Marvin S.
;    APPLICANT:  Franchini, Genoveffa
;    APPLICANT:  Markham, Phillip D.
;    APPLICANT:  Gallo, Robert C.
;    APPLICANT:  Lori, Franco C.
;    APPLICANT:  Popovic, Mikulas
;    APPLICANT:  Garnter, Suzanne
;    TITLE OF INVENTION:  MOLECULAR CLONES OF HIV-1 AND USES
;    TITLE OF INVENTION:  THEREOF
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN, DARBY & CUSHMAN
;      STREET:  Eleventh Floor, 1615 L. Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/022,835
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/599,491
;      FILING DATE:  17-OCT-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Scott, Watson T.
;      REGISTRATION NUMBER:  26,581
;      REFERENCE/DOCKET NUMBER:  WTS/5683/78507/SRL
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3000
;      TELEFAX:  (202) 822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9746 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  6243..8816
US-08-022-835-3

  Query Match             42.5%;  Score 1089.8;  DB 1;  Length 9746;
  Best Local Similarity   64.8%;  Pred. No. 8.9e-266;
  Matches 1632;  Conservative    0;  Mismatches  882;  Indels    3;  Gaps    1;

Qy       1 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 60
           |||||||   |    | ||  |||||  |||    |||||||||||||| |||||| |||
Db    2566 TTTCCCATTAGTCCTATTGAAACTGTACCAGTAAAATTAAAGCCAGGAATGGATGGCCCA 2625

Qy      61 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 120
           ||| | || || ||||||||  || |||||||  |||.  |||||||  |||||| |||  |
Db    2626 AAAGTTAAACAATGGCCATTGACAGAAGAAAAAATAAAAGCATTAATAGAAATTTGTACA 2685

Qy     121 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 180
           | |||||||||||  ||  |  |      ||  |   ||    ||||||||||  || ||
Db    2686 GAAATGGAAAAGGAAGGGAAAATTTCAAAAATTGGGCCTGAAAATCCATACAATACTCCA 2745

Qy     181 ACATTTGCTATAAAGAAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 240
           ||||||  ||||||||||| || |  |  ||||||||||    |  |||||||  || |||
Db    2746 GTATTTGCCATAAAGAAAAAAGACAGTACTAAATGGAGAAAATTAGTAGATTTCAGAGAA 2805

Qy     241 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 300
           || |||| |   |||||  |||||   ||||||  ||||||||||||||||||  ||||||||
Db    2806 CTTAATAAGAAAACTCAAGACTTCTGGGAAGTTCAATTAGGAATACCACATCCTGCAGGG 2865

Qy     301 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 360
           ||  |||| |  |||  | | ||||||||||| | ||||||||||||  ||   | ||
Db    2866 TTAAAAAAGAAAAAATCAGTAACAGTACTGGATGTGGGTGATGCATATTTTTCAGTTCCC 2925

Qy     361 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 420
           |||||  |||| || ||| |||| |||||  |||||  ||||     |||| |||| |
Db    2926 TTAGATAAAGACTTCAGGAAGTATACTGCATTTACCATACCTAGTATAAACAATGAAACA 2985

Qy     421 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 480
           |||||  |   ||||    || || || || || ||||||||||| || |||||||| ||
Db    2986 CCAGGGATTAGATATCAGTACAATGTGCTTCCACAGGGATGGAAAGGATCACCAGCAATA 3045

Qy     481 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 540
           ||||||   |  |||| | |  |  |||| || || || ||  |||||||| ||
Db    3046 TTCCAAAGTAGCATGACAAAAATCTTAGAGCCTTTTAGAAAACAAAATCCAGACATAGTT 3105

Qy     541 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 600
           |   || || |||||||| | |   |||    ||||     |     | | |||
Db    3106 ATCTATCAATACATGGATGATTTGTATGTAGGATCTGACTTAGAAATAGGGCAGCATAGA 3165

Qy     601 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 660
           | |    |   |    ||  ||  ||    || ||| | || ||||| |
Db    3166 GCAAAAATAGAGGAACTGAGACGACATCTGTTGAGGTGGGGATTTACCACACCAGACAAA 3225

Qy     661 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 720
           |||   || ||||| || ||||| |  |||||||| || ||| |    ||    ||||||
Db    3226 AAACATCAGAAAGAACCTCCATTCCTTTGGATGGGTTATGAACTCCATCCTGATAAATGG 3285

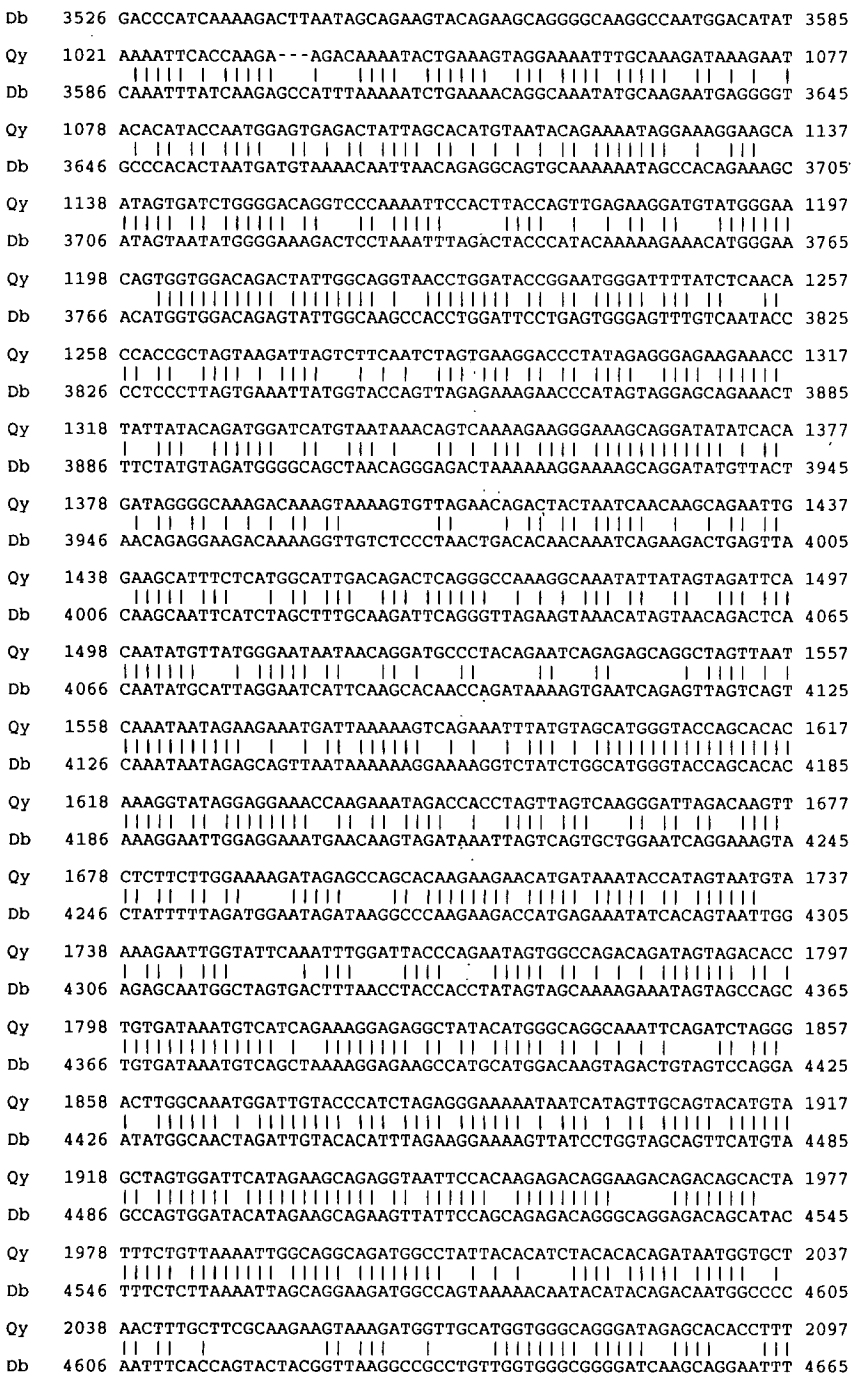
Qy     721 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 780
           |   | ||    |||| | ||||| ||| ||| | |||||| || ||||| |||||||||
Db    3286 ACAGTACAGCCTATAGTGCTGCCAGAAAAAGACAGCTGGACTGTCAATGACATACAGAAG 3345

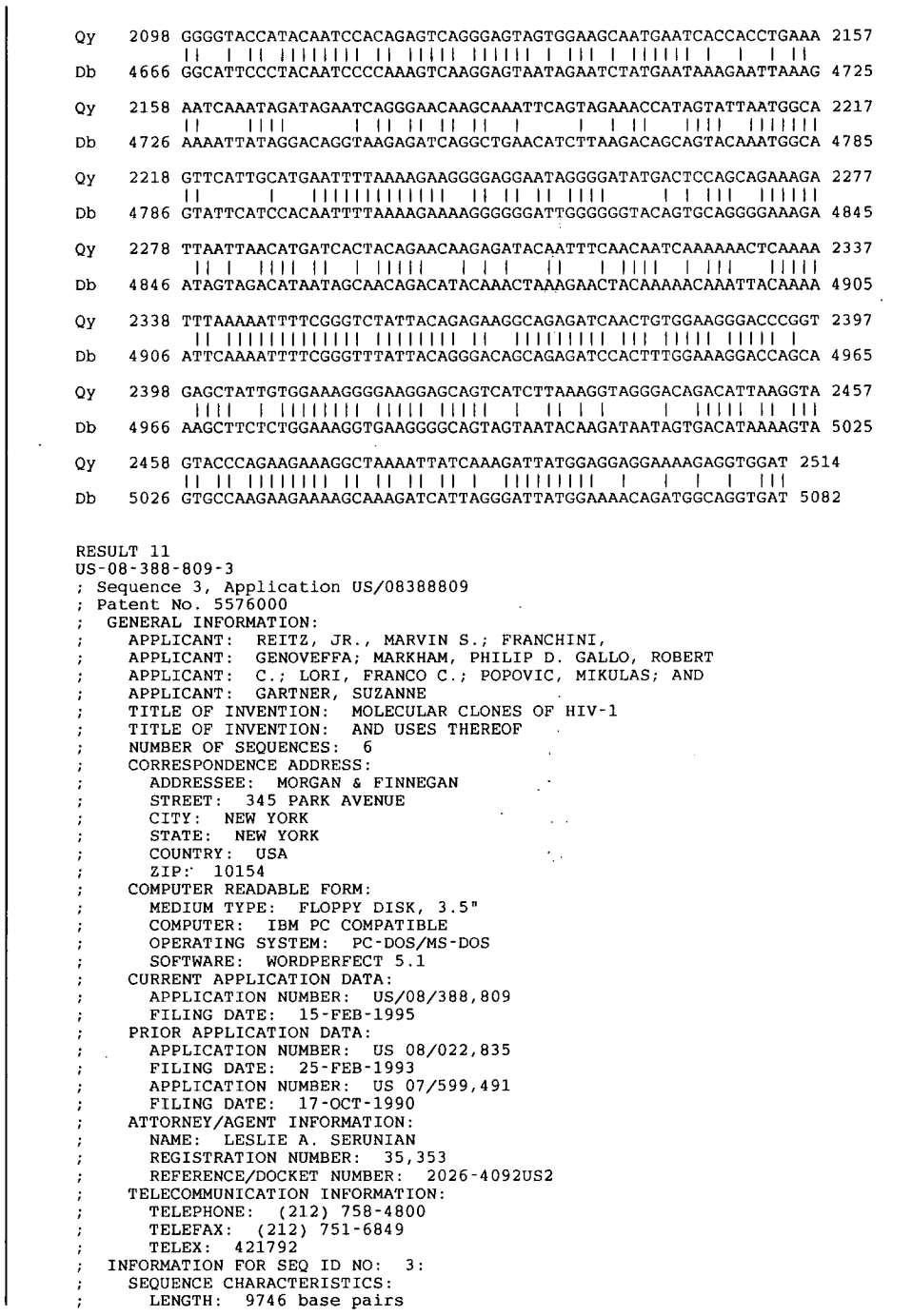
Qy     781 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 840
           ||||| |||  ||| |||||||||  |||||||||  |||| || |||   || ||  |
Db    3346 TTAGTGGGAAAATTGAATTGGGCAAGTCAAATTTACGCAGGGATTAAAGTAAAGCAATTA 3405

Qy     841 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 900
           ||||   |  |||||||||  |   | ||||||||| | | |     || ||  |  |||
Db    3406 TGTAAACTCCTTAGAGGAACCAAAGCACTAACAGAAGTAATACCACTAACAGAAGAAGCA 3465

Qy     901 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 960
           ||   ||||   |  ||||| |    |||||| |   |   |  | | |||   |||||
Db    3466 GAGCTAGAACTGGCAGAAAACAGGGAAATTCTAAAAGAACCAGTACATGGAGTGTATTAT 3525

Qy     961 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 1020
           |   |  ||    |||  |||    |||  |||     | |    || ||| | |||
```





```
RESULT 11
US-08-388-809-3
; Sequence 3, Application US/08388809
; Patent No. 5576000
;  GENERAL INFORMATION:
;    APPLICANT:  REITZ, JR., MARVIN S.; FRANCHINI,
;    APPLICANT:  GENOVEFFA; MARKHAM, PHILIP D. GALLO, ROBERT
;    APPLICANT:  C.; LORI, FRANCO C.; POPOVIC, MIKULAS; AND
;    APPLICANT:  GARTNER, SUZANNE
;    TITLE OF INVENTION:  MOLECULAR CLONES OF HIV-1
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN
;      STREET:  345 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK, 3.5"
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/388,809
;      FILING DATE:  15-FEB-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/022,835
;      FILING DATE:  25-FEB-1993
;      APPLICATION NUMBER:  US 07/599,491
;      FILING DATE:  17-OCT-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  LESLIE A. SERUNIAN
;      REGISTRATION NUMBER:  35,353
;      REFERENCE/DOCKET NUMBER:  2026-4092US2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 758-4800
;      TELEFAX:  (212) 751-6849
;      TELEX:  421792
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9746 base pairs
```

```
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  6243..8816
US-08-388-809-3

  Query Match             42.5%;  Score 1089.8;  DB 1;  Length 9746;
  Best Local Similarity   64.8%;  Pred. No. 8.9e-266;
  Matches 1632;  Conservative    0;  Mismatches  882;  Indels    3;  Gaps    1;

Qy       1 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 60
Db    2566 TTTCCCATTAGTCCTATTGAAACTGTACCAGTAAAATTAAAGCCAGGAATGGATGGCCCA 2625

Qy      61 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 120
Db    2626 AAAGTTAAACAATGGCCATTGACAGAAGAAAAAATAAAAGCATTAATAGAAATTTGTACA 2685

Qy     121 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 180
Db    2686 GAAATGGAAAAGGAAGGGAAAATTTCAAAAATTGGGCCTGAAAATCCATACAATACTCCA 2745

Qy     181 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 240
Db    2746 GTATTTGCCATAAAGAAAAAAGACAGTACTAAATGGAGAAAATTAGTAGATTTCAGAGAA 2805

Qy     241 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 300
Db    2806 CTTAATAAGAAAACTCAAGACTTCTGGGAAGTTCAATTAGGAATACCACATCCTGCAGGG 2865

Qy     301 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 360
Db    2866 TTAAAAAAGAAAAAATCAGTAACAGTACTGGATGTGGGTGATGCATATTTTTCAGTTCCC 2925

Qy     361 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 420
Db    2926 TTAGATAAAGACTTCAGGAAGTATACTGCATTTACCATACCTAGTATAAACAATGAAACA 2985

Qy     421 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 480
Db    2986 CCAGGGATTAGATATCAGTACAATGTGCTTCCACAGGGATGGAAAGGATCACCAGCAATA 3045

Qy     481 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 540
Db    3046 TTCCAAAGTAGCATGACAAAAATCTTAGAGCCTTTTAGAAAACAAAATCCAGACATAGTT 3105

Qy     541 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 600
Db    3106 ATCTATCAATACATGGATGATTTGTATGTAGGATCTGACTTAGAAATAGGGCAGCATAGA 3165

Qy     601 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 660
Db    3166 GCAAAAATAGAGGAACTGAGACGACATCTGTTGAGGTGGGGATTTACCACACCAGACAAA 3225

Qy     661 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 720
Db    3226 AAACATCAGAAAGAACCTCCATTCCTTTGGATGGGTTATGAACTCCATCCTGATAAATGG 3285

Qy     721 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 780
Db    3286 ACAGTACAGCCTATAGTGCTGCCAGAAAAAGACAGCTGGACTGTCAATGACATACAGAAG 3345

Qy     781 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 840
Db    3346 TTAGTGGGAAAATTGAATTGGGCAAGTCAAATTTACGCAGGGATTAAAGTAAAGCAATTA 3405

Qy     841 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 900
Db    3406 TGTAAACTCCTTAGAGGAACCAAAGCACTAACAGAAGTAATACCACTAACAGAAGAAGCA 3465

Qy     901 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 960

Db    3466 GAGCTAGAACTGGCAGAAAACAGGGAAATTCTAAAAGAACCAGTACATGGAGTGTATTAT 3525

Qy     961 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 1020
Db    3526 GACCCATCAAAAGACTTAATAGCAGAAGTACAGAAGCAGGGGCAAGGCCAATGGACATAT 3585

Qy    1021 AAAATTCACCAAGA---AGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAAT 1077
Db    3586 CAAATTTATCAAGAGCCATTTAAAAATCTGAAAACAGGCAAATATGCAAGAATGAGGGGT 3645

Qy    1078 ACACATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCA 1137
Db    3646 GCCCACACTAATGATGTAAAACAATTAACAGAGGCAGTGCAAAAAATAGCCACAGAAAGC 3705

Qy    1138 ATAGTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAA 1197
Db    3706 ATAGTAATATGGGGAAAGACTCCTAAATTTAGACTACCCATACAAAAAGAAACATGGGAA 3765

Qy    1198 CAGTGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACA 1257
Db    3766 ACATGGTGGACAGAGTATTGGCAAGCCACCTGGATTCCTGAGTGGGAGTTTGTCAATACC 3825

Qy    1258 CCACCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACC 1317
Db    3826 CCTCCCTTAGTGAAATTATGGTACCAGTTAGAGAAAGAACCCATAGTAGGAGCAGAAACT 3885

Qy    1318 TATTATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACA 1377
Db    3886 TTCTATGTAGATGGGGCAGCTAACAGGGAGACTAAAAAAGGAAAAGCAGGATATGTTACT 3945

Qy    1378 GATAGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTG 1437
Db    3946 AACAGAGGAAGACAAAAGGTTGTCTCCCTAACTGACACAACAAATCAGAAGACTGAGTTA 4005

Qy    1438 GAAGCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCA 1497
Db    4006 CAAGCAATTCATCTAGCTTTGCAAGATTCAGGGTTAGAAGTAAACATAGTAACAGACTCA 4065

Qy    1498 CAATATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAAT 1557
Db    4066 CAATATGCATTAGGAATCATTCAAGCACAACCAGATAAAAGTGAATCAGAGTTAGTCAGT 4125

Qy    1558 CAAATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACAC 1617
Db    4126 CAAATAATAGAGCAGTTAATAAAAAAGGAAAAGGTCTATCTGGCATGGGTACCAGCACAC 4185

Qy    1618 AAAGGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTT 1677
Db    4186 AAAGGAATTGGAGGAAATGAACAAGTAGATAAATTAGTCAGTGCTGGAATCAGGAAAGTA 4245

Qy    1678 CTCTTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTA 1737
Db    4246 CTATTTTTAGATGGAATAGATAAGGCCCAAGAAGACCATGAGAAATATCACAGTAATTGG 4305

Qy    1738 AAAGAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACC 1797
Db    4306 AGAGCAATGGCTAGTGACTTTAACCTACCACCTATAGTAGCAAAAGAAATAGTAGCCAGC 4365

Qy    1798 TGTGATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGG 1857
Db    4366 TGTGATAAATGTCAGCTAAAAGGAGAAGCCATGCATGGACAAGTAGACTGTAGTCCAGGA 4425

Qy    1858 ACTTGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTA 1917
Db    4426 ATATGGCAACTAGATTGTACACATTTAGAAGGAAAAGTTATCCTGGTAGCAGTTCATGTA 4485

Qy    1918 GCTAGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTA 1977
Db    4486 GCCAGTGGATACATAGAAGCAGAAGTTATTCCAGCAGAGACAGGGCAGGAGACAGCATAC 4545

Qy    1978 TTTCTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCT 2037
```

```
Db   4546 TTTCTCTTAAAATTAGCAGGAAGATGGCCAGTAAAAACAATACATACAGACAATGGCCCC 4605

Qy   2038 AACTTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTT 2097

Db   4606 AATTTCACCAGTACTACGGTTAAGGCCGCCTGTTGGTGGGCGGGGATCAAGCAGGAATTT 4665

Qy   2098 GGGGTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAA 2157

Db   4666 GGCATTCCCTACAATCCCCAAAGTCAAGGAGTAATAGAATCTATGAATAAAGAATTAAAG 4725

Qy   2158 AATCAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCA 2217

Db   4726 AAAATTATAGGACAGGTAAGAGATCAGGCTGAACATCTTAAGACAGCAGTACAAATGGCA 4785

Qy   2218 GTTCATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGA 2277

Db   4786 GTATTCATCCACAATTTTAAAAGAAAAGGGGGGATTGGGGGGTACAGTGCAGGGGAAAGA 4845

Qy   2278 TTAATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAA 2337

Db   4846 ATAGTAGACATAATAGCAACAGACATACAAACTAAAGAACTACAAAAACAAATTACAAAA 4905

Qy   2338 TTTAAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGT 2397

Db   4906 ATTCAAAATTTTCGGGTTTATTACAGGGACAGCAGAGATCCACTTTGGAAAGGACCAGCA 4965

Qy   2398 GAGCTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTA 2457

Db   4966 AAGCTTCTCTGGAAAGGTGAAGGGGCAGTAGTAATACAAGATAATAGTGACATAAAAGTA 5025

Qy   2458 GTACCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGAT 2514

Db   5026 GTGCCAAGAAGAAAAGCAAAGATCATTAGGGATTATGGAAAACAGATGGCAGGTGAT 5082
```

```
RESULT 12
US-08-647-714-3
; Sequence 3, Application US/08647714
; Patent No. 5869313
;  GENERAL INFORMATION:
;    APPLICANT:  REITZ, JR., MARVIN S.; FRANCHINI,
;    APPLICANT:  GENOVEFFA; MARKHAM, PHILIP D. GALLO, ROBERT
;    APPLICANT:  C.; LORI, FRANCO C.; POPOVIC, MIKULAS; AND
;    APPLICANT:  GARTNER, SUZANNE
;    TITLE OF INVENTION:  MOLECULAR CLONES OF HIV-1
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN
;      STREET:  345 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK, 3.5"
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/647,714
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/388,809
;      FILING DATE:  15-FEB-1995
;      APPLICATION NUMBER:  US 08/022,835
;      FILING DATE:  25-FEB-1993
;      APPLICATION NUMBER:  US 07/599,491
;      FILING DATE:  17-OCT-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  LESLIE A. SERUNIAN
;      REGISTRATION NUMBER:  35,353
;      REFERENCE/DOCKET NUMBER:  2026-4092US2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 758-4800
;      TELEFAX:  (212) 751-6849
;      TELEX:  421792
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9746 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  6243..8816
US-08-647-714-3

  Query Match             42.5%;  Score 1089.8;  DB 2;  Length 9746;
  Best Local Similarity   64.8%;  Pred. No. 8.9e-266;
  Matches 1632;  Conservative    0;  Mismatches  882;  Indels    3;  Gaps    1;

Qy       1 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 60

Db    2566 TTTCCCATTAGTCCTATTGAAACTGTACCAGTAAAATTAAAGCCAGGAATGGATGGCCCA 2625

Qy      61 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 120

Db    2626 AAAGTTAAACAATGGCCATTGACAGAAGAAAAAATAAAAGCATTAATAGAAATTTGTACA 2685

Qy     121 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 180

Db    2686 GAAATGGAAAAGGAAGGGAAAATTTCAAAAATTGGGCCTGAAAATCCATACAATACTCCA 2745

Qy     181 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 240

Db    2746 GTATTTGCCATAAAGAAAAAAGACAGTACTAAATGGAGAAAATTAGTAGATTTCAGAGAA 2805

Qy     241 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 300

Db    2806 CTTAATAAGAAAACTCAAGACTTCTGGGAAGTTCAATTAGGAATACCACATCCTGCAGGG 2865

Qy     301 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 360

Db    2866 TTAAAAAAGAAAAAATCAGTAACAGTACTGGATGTGGGTGATGCATATTTTTCAGTTCCC 2925

Qy     361 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 420

Db    2926 TTAGATAAAGACTTCAGGAAGTATACTGCATTTACCATACCTAGTATAAACAATGAAACA 2985

Qy     421 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 480

Db    2986 CCAGGGATTAGATATCAGTACAATGTGCTTCCACAGGGATGGAAAGGATCACCAGCAATA 3045

Qy     481 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 540

Db    3046 TTCCAAAGTAGCATGACAAAAATCTTAGAGCCTTTTAGAAAACAAAATCCAGACATAGTT 3105

Qy     541 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 600

Db    3106 ATCTATCAATACATGGATGATTTGTATGTAGGATCTGACTTAGAAATAGGGCAGCATAGA 3165

Qy     601 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 660

Db    3166 GCAAAAATAGAGGAACTGAGACGACATCTGTTGAGGTGGGGATTTACCACACCAGACAAA 3225

Qy     661 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 720

Db    3226 AAACATCAGAAAGAACCTCCATTCCTTTGGATGGGTTATGAACTCCATCCTGATAAATGG 3285

Qy     721 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 780

Db    3286 ACAGTACAGCCTATAGTGCTGCCAGAAAAAGACAGCTGGACTGTCAATGACATACAGAAG 3345

Qy     781 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 840
```

```
Db   3346 TTAGTGGGAAAATTGAATTGGGCAAGTCAAATTTACGCAGGGATTAAAGTAAAGCAATTA 3405

Qy    841 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 900

Db   3406 TGTAAACTCCTTAGAGGAACCAAAGCACTAACAGAAGTAATACCACTAACAGAAGAAGCA 3465

Qy    901 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 960

Db   3466 GAGCTAGAACTGGCAGAAAACAGGGAAATTCTAAAAGAACCAGTACATGGAGTGTATTAT 3525

Qy    961 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 1020

Db   3526 GACCCATCAAAAGACTTAATAGCAGAAGTACAGAAGCAGGGGCAAGGCCAATGGACATAT 3585

Qy   1021 AAAATTCACCAAGA---AGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAAT 1077

Db   3586 CAAATTTATCAAGAGCCATTTAAAAATCTGAAAACAGGCAAATATGCAAGAATGAGGGGT 3645

Qy   1078 ACACATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCA 1137

Db   3646 GCCCACACTAATGATGTAAAACAATTAACAGAGGCAGTGCAAAAAATAGCCACAGAAAGC 3705

Qy   1138 ATAGTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAA 1197

Db   3706 ATAGTAATATGGGGAAAGACTCCTAAATTTAGACTACCCATACAAAAAGAAACATGGGAA 3765

Qy   1198 CAGTGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACA 1257

Db   3766 ACATGGTGGACAGAGTATTGGCAAGCCACCTGGATTCCTGAGTGGGAGTTTGTCAATACC 3825

Qy   1258 CCACCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACC 1317

Db   3826 CCTCCCTTAGTGAAATTATGGTACCAGTTAGAGAAAGAACCCATAGTAGGAGCAGAAACT 3885

Qy   1318 TATTATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACA 1377

Db   3886 TTCTATGTAGATGGGGCAGCTAACAGGGAGACTAAAAAAGGAAAAGCAGGATATGTTACT 3945

Qy   1378 GATAGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTG 1437

Db   3946 AACAGAGGAAGACAAAAGGTTGTCTCCCTAACTGACACAACAAATCAGAAGACTGAGTTA 4005

Qy   1438 GAAGCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCA 1497

Db   4006 CAAGCAATTCATCTAGCTTTGCAAGATTCAGGGTTAGAAGTAAACATAGTAACAGACTCA 4065

Qy   1498 CAATATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAAT 1557

Db   4066 CAATATGCATTAGGAATCATTCAAGCACAACCAGATAAAAGTGAATCAGAGTTAGTCAGT 4125

Qy   1558 CAAATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACAC 1617

Db   4126 CAAATAATAGAGCAGTTAATAAAAAAGGAAAAGGTCTATCTGGCATGGGTACCAGCACAC 4185

Qy   1618 AAAGGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTT 1677

Db   4186 AAAGGAATTGGAGGAAATGAACAAGTAGATAAATTAGTCAGTGCTGGAATCAGGAAAGTA 4245

Qy   1678 CTCTTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTA 1737

Db   4246 CTATTTTTAGATGGAATAGATAAGGCCCAAGAAGACCATGAGAAATATCACAGTAATTGG 4305

Qy   1738 AAAGAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACC 1797

Db   4306 AGAGCAATGGCTAGTGACTTTAACCTACCACCTATAGTAGCAAAAGAAATAGTAGCCAGC 4365

Qy   1798 TGTGATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGG 1857

Db   4366 TGTGATAAATGTCAGCTAAAAGGAGAAGCCATGCATGGACAAGTAGACTGTAGTCCAGGA 4425

Qy   1858 ACTTGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTA 1917

Db   4426 ATATGGCAACTAGATTGTACACATTTAGAAGGAAAAGTTATCCTGGTAGCAGTTCATGTA 4485

Qy   1918 GCTAGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTA 1977

Db   4486 GCCAGTGGATACATAGAAGCAGAAGTTATTCCAGCAGAGACAGGGCAGGAGACAGCATAC 4545

Qy   1978 TTTCTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCT 2037

Db   4546 TTTCTCTTAAAATTAGCAGGAAGATGGCCAGTAAAAACAATACATACAGACAATGGCCCC 4605

Qy   2038 AACTTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTT 2097

Db   4606 AATTTCACCAGTACTACGGTTAAGGCCGCCTGTTGGTGGGCGGGGATCAAGCAGGAATTT 4665

Qy   2098 GGGGTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAA 2157

Db   4666 GGCATTCCCTACAATCCCCAAAGTCAAGGAGTAATAGAATCTATGAATAAAGAATTAAAG 4725

Qy   2158 AATCAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCA 2217

Db   4726 AAAATTATAGGACAGGTAAGAGATCAGGCTGAACATCTTAAGACAGCAGTACAAATGGCA 4785

Qy   2218 GTTCATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGA 2277

Db   4786 GTATTCATCCACAATTTTAAAAGAAAAGGGGGGATTGGGGGGTACAGTGCAGGGGAAAGA 4845

Qy   2278 TTAATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAA 2337

Db   4846 ATAGTAGACATAATAGCAACAGACATACAAACTAAAGAACTACAAAAACAAATTACAAAA 4905

Qy   2338 TTTAAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGT 2397

Db   4906 ATTCAAAATTTTCGGGTTTATTACAGGGACAGCAGAGATCCACTTTGGAAAGGACCAGCA 4965

Qy   2398 GAGCTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTA 2457

Db   4966 AAGCTTCTCTGGAAAGGTGAAGGGGCAGTAGTAATACAAGATAATAGTGACATAAAAGTA 5025

Qy   2458 GTACCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGAT 2514

Db   5026 GTGCCAAGAAGAAAAGCAAAGATCATTAGGGATTATGGAAAACAGATGGCAGGTGAT 5082
```

RESULT 13
US-08-463-210-5
; Sequence 5, Application US/08463210
; Patent No. 6001977
; GENERAL INFORMATION:
; APPLICANT: CHANG, Nancy T.
; APPLICANT: GALLO, Robert C.
; APPLICANT: WONG-STAAL, Flossie
; TITLE OF INVENTION: CLONING AND EXPRESSION OF HTLV-III DNA
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morgan & Finnegan, L.L.P.
; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154-0053
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,210
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/693,866
; FILING DATE: 23-JAN-1985
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/659,339

```
;      FILING DATE:  10-OCT-1984
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Serunian, Leslie A.
;      REGISTRATION NUMBER:  35,353
;      REFERENCE/DOCKET NUMBER:  2026-4193US2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 758-4800
;      TELEFAX:  (212) 751-6849
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5362 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  HTLV-III
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..5362
;      OTHER INFORMATION:  /standard_name= "Clone BH5"
;      OTHER INFORMATION:  /note= "Corresponds to nucleotide positions 222 to
;      OTHER INFORMATION:  5585 in figure 3 of US 06/693,866 (parent)"
US-08-463-210-5

  Query Match             42.3%;  Score 1083.4;  DB 3;  Length 5362;
  Best Local Similarity   64.7%;  Pred. No. 2.9e-264;
  Matches 1628;  Conservative    0;  Mismatches  886;  Indels    3;  Gaps    1;

Qy       1 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 60
Db    1904 TTTCCCATTAGTCCTATTGAAACTGTACCAGTAAAATTAAAGCCAGGAATGGATGGCCCA 1963

Qy      61 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 120
Db    1964 AAAGTTAAACAATGGCCATTGACAGAAGAAAAAATAAAAGCATTAGTAGAAATTTGTACA 2023

Qy     121 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 180
Db    2024 GAAATGGAAAAGGAAGGGAAAATTTCAAAAATTGGGCCTGAAAATCCATACAATACTCCA 2083

Qy     181 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 240
Db    2084 GTATTTGCCATAAAGAAAAAAGACAGTACTAAATGGAGAAAATTAGTAGATTTCAGAGAA 2143

Qy     241 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 300
Db    2144 CTTAATAGGAGAACTCAAGACTTCTGGGAAGTTCAATTGGGAATACCACATCCCGCAGGG 2203

Qy     301 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 360
Db    2204 TTAAAAAAGAAAAAATCAGTAACAGTACTGGATGTGGGTGATGCATATTTTTCAGTTCCC 2263

Qy     361 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 420
Db    2264 TTAGATGAAGACTTCAGGAAGTATACTGCATTTACCATACCTAGTATAAATAATGAGACA 2323

Qy     421 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 480
Db    2324 CCAGGGAGTGGATATCAGTACAATGTGCTTCCACAGGGATGGAAAGGATCACCAGCAATA 2383

Qy     481 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 540
Db    2384 TTCCAAAGTAGCATGACAAAAATCTTAGAGCCTTTTAGAAAACAAAATCCAGACATAGTT 2443

Qy     541 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 600
Db    2444 ATTTATCAATACATGGATGATTTGTATGTAGGATCTGACTTAGAAATAGGGCAGCATAGA 2503

Qy     601 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 660
Db    2504 ACAAAAATAGAGGAGCTGAGACAACATCTGTTGAGGTGGGGATTTACCACACCAGACAAA 2563

Qy     661 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 720
Db    2564 AAACATCAGAAAGAACCTCCATTCCTTTGGATGGGTTATGAACTCCATCCTGATAAATGG 2623

Qy     721 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 780
Db    2624 ACGATACAGCCTATAGTGCTGCCAGAAAAAGACAGCTGGACTGTCAATGACATACAGAAG 2683

Qy     781 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 840
Db    2684 TTAGTGGGAAAATTGAATTGGGCAAGTCAGATTTATCCAGGGATTAAAGTAAGGCAATTA 2743

Qy     841 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 900
Db    2744 TGTAAACTCCTTAGAGGAACCAAAGCACTAACAGAAGTAATACCACTAACAGAAGAAGCA 2803

Qy     901 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 960
Db    2804 GAGCTAGAACTGGCAGAAAACAGAGAGATTCTAAAAGAACCAGTACATGGAGTGTATTAT 2863

Qy     961 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 1020
Db    2864 GACCCATCAAAAGACTTAATAGCAGAAATACAGAAGCAGGGGCAAGGCCAATGGACATAT 2923

Qy    1021 AAAATTCACCAAGA---AGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAAT 1077
Db    2924 CAAATTTATCAAGAGCCATTTAAAAATCTGAAAACAGGAAAATATGCAAGAATGAGGGGT 2983

Qy    1078 ACACATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCA 1137
Db    2984 GCCCACACTAATGATGTAAAACAATTAACAGAGGCAGTGCAAAAAATAACCACAGAAAGC 3043

Qy    1138 ATAGTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAA 1197
Db    3044 ATAGTAATATGGGGAAAGACTCCTAAATTTAAACTACCCATACAAAAAGAAACATGGGAA 3103

Qy    1198 CAGTGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACA 1257
Db    3104 ACATGGTGGACAGAGTATTGGCAAGCCACCTGGATTCCTGAGTGGGAGTTTGTTAATACC 3163

Qy    1258 CCACCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACC 1317
Db    3164 CCTCCTTTAGTGAAATTATGGTACCAGTTAGAGAAAGAACCCATAGTAGGAGCAGAAACC 3223

Qy    1318 TATTATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACA 1377
Db    3224 TTCTATGTAGATGGGGCAGCTAGCAGGGAGACTAAATTAGGAAAAGCAGGATATGTTACT 3283

Qy    1378 GATAGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTG 1437
Db    3284 AATAGAGGAAGACAAAAAGTTGTCACCCTAACTCACACAACAAATCAGAAGACTGAATTA 3343

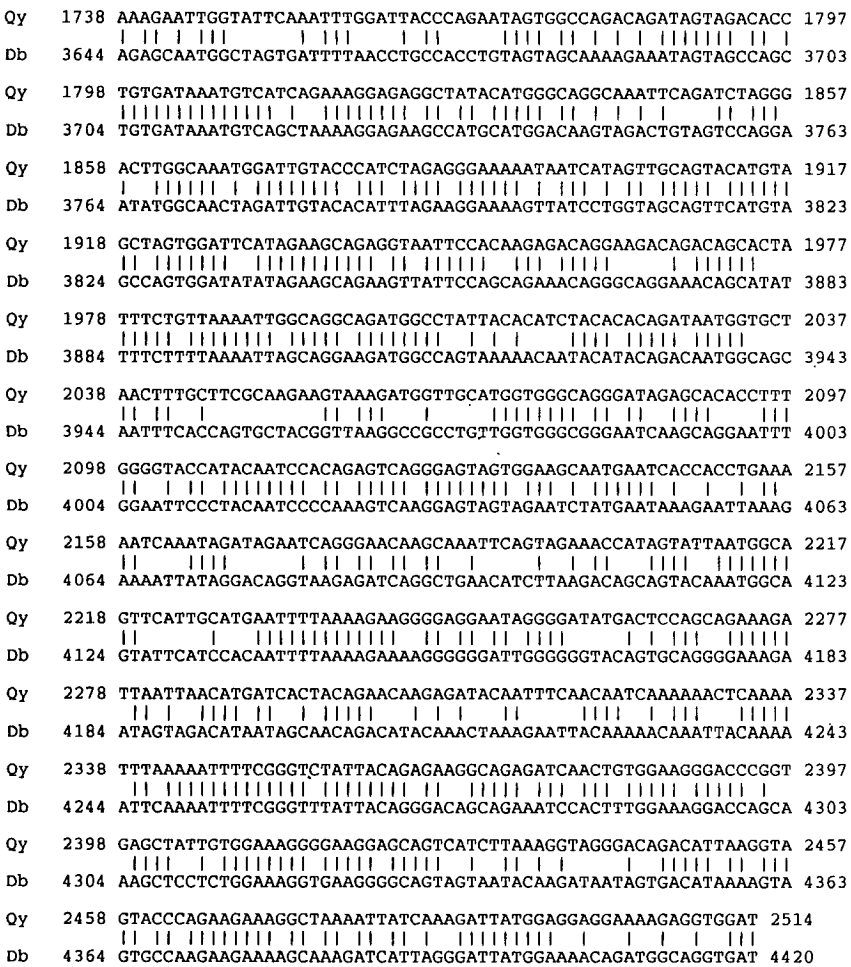
Qy    1438 GAAGCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCA 1497
Db    3344 CAAGCAATTCATCTAGCTTTGCAGGATTCGGGATTAGAAGTAAATATAGTAACAGACTCA 3403

Qy    1498 CAATATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAAT 1557
Db    3404 CAATATGCATTAGGAATCATTCAAGCACAACCAGATAAAAGTGAATCAGAGTTAGTCAAT 3463

Qy    1558 CAAATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACAC 1617
Db    3464 CAAATAATAGAGCAGTTAATAAAAAAGGAAAAGGTCTATCTGGCATGGGTACCAGCACAC 3523

Qy    1618 AAAGGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTT 1677
Db    3524 AAAGGAATTGGAGGAAATGAACAAGTAGATAAATTAGTCAGTGCTGGAATCAGGAAAATA 3583

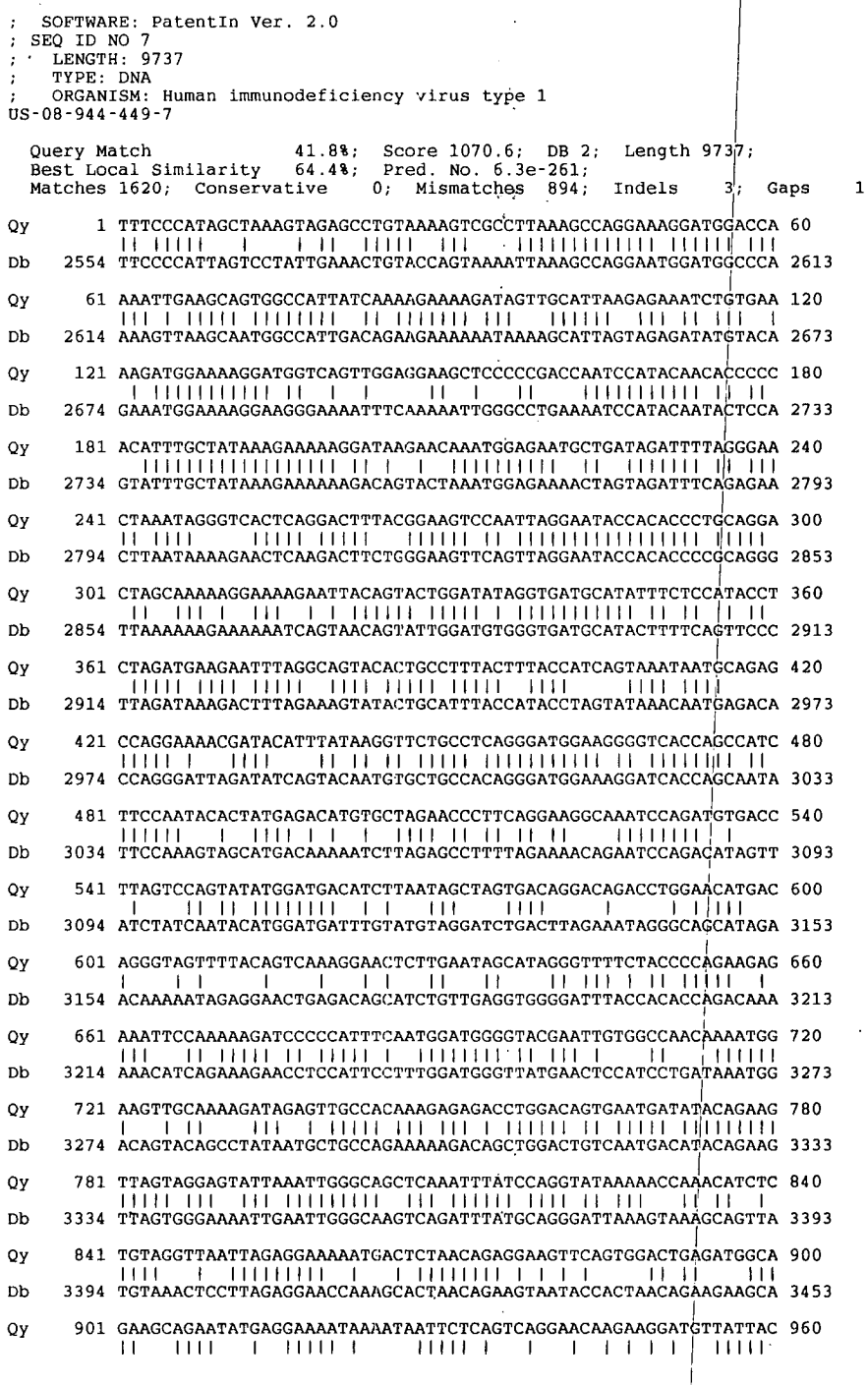
Qy    1678 CTCTTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTA 1737
Db    3584 CTATTTTTAGATGGAATAGATAAGGCCCAAGAAGAACATGAGAAATATCACAGTAATTGG 3643
```



RESULT 14
US-08-944-449-7
; Sequence 7, Application US/08944449
; Patent No. 5985613
; GENERAL INFORMATION:
; APPLICANT: KURTH, REINHARD
; APPLICANT: BAIER, MICHAEL
; APPLICANT: METZNER, KARIN
; APPLICANT: WERNER, ALBRECHT
; TITLE OF INVENTION: Use of an "immunodeficiency-virus suppressing
; TITLE OF INVENTION: lymphokine (ISL)" to inhibit the replication of
; TITLE OF INVENTION: viruses, particularly of retroviruses
; FILE REFERENCE: 8341-7065
; CURRENT APPLICATION NUMBER: US/08/944,449
; CURRENT FILING DATE: 1997-10-06
; EARLIER APPLICATION NUMBER: EP 95113013.2
; EARLIER FILING DATE: 1995-08-18
; EARLIER APPLICATION NUMBER: DE 195 13 152.5
; EARLIER FILING DATE: 1995-04-07
; NUMBER OF SEQ ID NOS: 8

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 9737
; TYPE: DNA
; ORGANISM: Human immunodeficiency virus type 1
US-08-944-449-7

Query Match 41.8%; Score 1070.6; DB 2; Length 9737;
Best Local Similarity 64.4%; Pred. No. 6.3e-261;
Matches 1620; Conservative 0; Mismatches 894; Indels 3; Gaps 1;



```
Db   3454 GAGCTAGAACTGGCAGAAAACAGGGAGATTCTAAAAGAACCAGTACATGAAGTATATTAT 3513

Qy    961 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 1020

Db   3514 GACCCATCAAAAGACTTAGTAGCAGAAATACAGAAGCAGGGGCAAGGCCAATGGACATAT 3573

Qy   1021 AAAATTCACCAAGA---AGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAAT 1077

Db   3574 CAAATTTATCAAGAGCCATTTAAAAATCTGAAAACAGGAAAGTATGCAAGGATGAGGGGT 3633

Qy   1078 ACACATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCA 1137

Db   3634 GCCCACACTAATGATGTAAAACAGTTAACAGAGGCAGTGCAAAAAGTATCCACAGAAAGC 3693

Qy   1138 ATAGTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAA 1197

Db   3694 ATAGTAATATGGGGAAAGATTCCTAAATTTAAACTACCCATACAAAAGGAAACATGGGAA 3753

Qy   1198 CAGTGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACA 1257

Db   3754 GCATGGTGGATGGAGTATTGGCAAGCTACCTGGATTCCTGAGTGGGAGTTTGTCAATACC 3813

Qy   1258 CCACCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACC 1317

Db   3814 CCTCCCTTAGTGAAATTATGGTACCAGTTAGAGAAAGAACCCATAGTAGGAGCAGAAACT 3873

Qy   1318 TATTATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACA 1377

Db   3874 TTCTATGTAGATGGGGCAGCTAATAGGGAGACTAAATTAGGAAAAGCAGGATATGTTACT 3933

Qy   1378 GATAGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTG 1437

Db   3934 GACAGAGGAAGACAAAAAGTTGTCTCCATAGCTGACACAACAAATCAGAAGACTGAATTA 3993

Qy   1438 GAAGCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCA 1497

Db   3994 CAAGCAATTCATCTAGCTTTGCAGGATTCGGGATTAGAAGTAAACATAGTAACAGACTCA 4053

Qy   1498 CAATATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAAT 1557

Db   4054 CAATATGCATTAGGAATCATTCAAGCACAACCAGATAAGAGTGAATCAGAGTTAGTCAGT 4113

Qy   1558 CAAATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACAC 1617

Db   4114 CAAATAATAGAGCAGTTAATAAAAAAGGAAAAGGTCTACCTGGCATGGGTACCAGCACAC 4173

Qy   1618 AAAGGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTT 1677

Db   4174 AAAGGAATTGGAGGAAATGAACAAGTAGATAAATTAGTCAGTGCTGGAATCAGGAAAGTA 4233

Qy   1678 CTCTTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTA 1737

Db   4234 CTATTTTTGAATGGAATAGATAAGGCCCAAGAAGAACATGAGAAATATCACAGTAATTGG 4293

Qy   1738 AAAGAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACC 1797

Db   4294 AGAGCAATGGCTAGTGATTTTAACCTGCCACCTGTAGTAGCAAAAGAAATAGTAGCCAGC 4353

Qy   1798 TGTGATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGG 1857

Db   4354 TGTGATAAATGTCAGCTAAAAGGAGAAGCCATGCATGGACAAGTAGACTGTAGTCCAGGA 4413

Qy   1858 ACTTGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTA 1917

Db   4414 ATATGGCAACTAGATTGTACACATCTAGAAGGAAAAATTATCCTGGTAGCAGTTCATGTA 4473

Qy   1918 GCTAGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTA 1977

Db   4474 GCCAGTGGATATATAGAAGCAGAAGTTATTCCAGCAGAGACAGGGCAGGAAACAGCATAT 4533

Qy   1978 TTTCTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCT 2037

Db   4534 TTTCTCTTAAAATTAGCAGGAAGATGGCCAGTAAAAACAATACATACAGACAATGGCAGC 4593

Qy   2038 AACTTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTT 2097

Db   4594 AATTTCACCAGTACTACGGTTAAGGCCGCCTGTTGGTGGGCAGGGATCAAGCAGGAATTT 4653

Qy   2098 GGGGTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAA 2157

Db   4654 GGCATTCCCTACAATCCCCAAAGTCAAGGAGTAGTAGAATCTATGAATAATGAATTAAAG 4713

Qy   2158 AATCAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCA 2217

Db   4714 AAAATTATAGGACAGGTAAGAGATCAGGCTGAACACCTTAAGACAGCAGTACAAATGGCA 4773

Qy   2218 GTTCATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGA 2277

Db   4774 GTATTCATCCACAATTTTAAAAGAAAAGGGGGGATTGGGGGATACAGTGCAGGGGAAAGA 4833

Qy   2278 TTAATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAA 2337

Db   4834 ATAGTAGACATAATAGCAACAGACATACAAACTAAAGAACTACAAAAGCAAATTACAAAA 4893

Qy   2338 TTTAAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGT 2397

Db   4894 ATTCAAAATTTTCGGGTTTATTACAGGGACAACAAAGATCCCCTTTGGAAAGGACCAGCA 4953

Qy   2398 GAGCTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTA 2457

Db   4954 AAGCTTCTCTGGAAAGGTGAAGGGGCAGTAGTAATACAAGATAATAGTGACATAAAAGTA 5013

Qy   2458 GTACCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGAT 2514

Db   5014 GTGCCAAGAAGAAAAGCAAAAATCATTAGGGATTATGGAAAACAGATGGCAGGTGAT 5070
```

RESULT 15
US-09-353-362-7
; Sequence 7, Application US/09353362
; Patent No. 6383739
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Use of an "immunodeficiency-virus suppressing
; TITLE OF INVENTION: lymphokine (ISL)" to inhibit the replication of viruses,
; TITLE OF INVENTION: in particular of retroviruses
; NUMBER OF SEQUENCES: 8
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/353,362
; FILING DATE: 15-JUL-1999
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 195 13 152.5
; FILING DATE: 07-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 95113013.2
; FILING DATE: 18-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KLESNER, Sharon N.
; REGISTRATION NUMBER: 36,335
; REFERENCE/DOCKET NUMBER: P8341-9012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 638-5000
; TELEFAX: (202) 638-4810
; TELEX:
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9737 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear

```
;    MOLECULE TYPE:  DNA (genomic)
US-09-353-362-7

  Query Match             41.8%;  Score 1070.6;  DB 4;  Length 9737;
  Best Local Similarity   64.4%;  Pred. No. 6.3e-261;
  Matches 1620;  Conservative    0;  Mismatches  894;  Indels    3;  Gaps    1;

Qy       1 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGGCAGGAAAGGATGGACCA 60
           || |||||   |    | ||  |||||  |||   ||||||||||||| |||||| |||
Db    2554 TTCCCCATTAGTCCTATTGAAACTGTACCAGTAAAATTAAAGCCAGGAATGGATGGCCCA 2613

Qy      61 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 120
           ||| | ||||| ||||||||  || |||||||.|||   ||||||  ||| || |||  |
Db    2614 AAAGTTAAGCAATGGCCATTGACAGAAGAAAAAATAAAAGCATTAGTAGAGATATGTACA 2673

Qy     121 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 180
           | ||||||||||| ||  |  |      ||  |  ||    ||||||||||| || ||
Db    2674 GAAATGGAAAAGGAAGGGAAAATTTCAAAAATTGGGCCTGAAAATCCATACAATACTCCA 2733

Qy     181 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 240
           ||||||||||||||||||| || |  |  ||||||||||  ||  ||||||| || |||
Db    2734 GTATTTGCTATAAAGAAAAAAGACAGTACTAAATGGAGAAAACTAGTAGATTTCAGAGAA 2793

Qy     241 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 300
           || ||||     ||||| |||||   |||||| || ||||||||||||||||| |||||
Db    2794 CTTAATAAAAGAACTCAAGACTTCTGGGAAGTTCAGTTAGGAATACCACACCCCGCAGGG 2853

Qy     301 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 360
           ||  ||| |  |||  | | ||||||  ||||| | ||||||||||  || ||  |  ||
Db    2854 TTAAAAAAGAAAAAATCAGTAACAGTATTGGATGTGGGTGATGCATACTTTTCAGTTCCC 2913

Qy     361 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 420
           |||||  ||||  |||||  ||||  |||||  |||||  ||||      ||||  ||||
Db    2914 TTAGATAAAGACTTTAGAAAGTATACTGCATTTACCATACCTAGTATAAACAATGAGACA 2973

Qy     421 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 480
           |||||  |   ||||    || || || |||||  |||||||||||  || ||||||||  ||
Db    2974 CCAGGGATTAGATATCAGTACAATGTGCTGCCACAGGGATGGAAAGGATCACCAGCAATA 3033

Qy     481 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 540
           ||||||   |  |||| | |   |||| || || || ||    ||||||||  |
Db    3034 TTCCAAAGTAGCATGACAAAAATCTTAGAGCCTTTTAGAAAACAGAATCCAGACATAGTT 3093

Qy     541 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 600
           |    || || ||||||||  | |   |||   ||||      |    | | |||
Db    3094 ATCTATCAATACATGGATGATTTGTATGTAGGATCTGACTTAGAAATAGGGCAGCATAGA 3153

Qy     601 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 660
           |    | |      |    |  | |   ||   ||    || ||| | || |||||  |
Db    3154 ACAAAAATAGAGGAACTGAGACAGCATCTGTTGAGGTGGGGATTTACCACACCAGACAAA 3213

Qy     661 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 720
           |||   || ||||| || ||||| |  |||||||| || ||| |    ||    ||||||
Db    3214 AAACATCAGAAAGAACCTCCATTCCTTTGGATGGGTTATGAACTCCATCCTGATAAATGG 3273

Qy     721 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 780
           |   | ||    |||  | ||||| ||| ||| | ||||||  || |||||  |||||||||
Db    3274 ACAGTACAGCCTATAATGCTGCCAGAAAAAGACAGCTGGACTGTCAATGACATACAGAAG 3333

Qy     781 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 840
           ||||| |||  ||| |||||||||  ||| |||||| |||| || |||   || ||  |
Db    3334 TTAGTGGGAAAATTGAATTGGGCAAGTCAGATTTATGCAGGGATTAAAGTAAAGCAGTTA 3393

Qy     841 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 900
           ||||   |  |||||||||  |   | |||||||| | | | |     || ||    |||
Db    3394 TGTAAACTCCTTAGAGGAACCAAAGCACTAACAGAAGTAATACCACTAACAGAAGAAGCA 3453

Qy     901 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 960
           ||  ||||   | ||||| |     ||||| |   |   |  | | | |   |||||
Db    3454 GAGCTAGAACTGGCAGAAAACAGGGAGATTCTAAAAGAACCAGTACATGAAGTATATTAT 3513

Qy     961 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 1020
             |   |   ||    |||| |||     ||   |||     | |    || ||| | |||

Db    3514 GACCCATCAAAAGACTTAGTAGCAGAAATACAGAAGCAGGGGCAAGGCCAATGGACATAT 3573

Qy    1021 AAAATTCACCAAGA---AGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAAT 1077
           ||||| | |||||    |   ||||  ||||||  |||||| | ||||| ||| | ||  |
Db    3574 CAAATTTATCAAGAGCCATTTAAAAATCTGAAAACAGGAAAGTATGCAAGGATGAGGGGT 3633

Qy    1078 ACACATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCA 1137
           | || || ||||  || | ||  ||| || | | | | || ||| ||    |  |||
Db    3634 GCCCACACTAATGATGTAAAACAGTTAACAGAGGCAGTGCAAAAAGTATCCACAGAAAGC 3693

Qy    1138 ATAGTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAA 1197
           ||||| || |||||| || | || |||||  |  ||||  |  | |||||   ||||||||
Db    3694 ATAGTAATATGGGGAAAGATTCCTAAATTTAAACTACCCATACAAAAGGAAACATGGGAA 3753

Qy    1198 CAGTGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACA 1257
           |||||||  || ||||||||  |  ||||||||.|| || ||||| ||| ||  | ||
Db    3754 GCATGGTGGATGGAGTATTGGCAAGCTACCTGGATTCCTGAGTGGGAGTTTGTCAATACC 3813

Qy    1258 CCACCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACC 1317
           || ||  |||| | ||||   | | |  ||| ||| || || ||||  |||| ||||||
Db    3814 CCTCCCTTAGTGAAATTATGGTACCAGTTAGAGAAAGAACCCATAGTAGGAGCAGAAACT 3873

Qy    1318 TATTATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACA 1377
           |  |||  ||||||  ||  |||||   || | |||  ||| ||||||||||||| | ||
Db    3874 TTCTATGTAGATGGGGCAGCTAATAGGGAGACTAAATTAGGAAAAGCAGGATATGTTACT 3933

Qy    1378 GATAGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTG 1437
           || || || | | | |||||         |||   | || || |||||  |  | |||||
Db    3934 GACAGAGGAAGACAAAAAGTTGTCTCCATAGCTGACACAACAAATCAGAAGACTGAATTA 3993

Qy    1438 GAAGCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCA 1497
           |||||  |||   | || ||  || ||   |  | | ||| ||  ||  ||| |||
Db    3994 CAAGCAATTCATCTAGCTTTGCAGGATTCGGGATTAGAAGTAAACATAGTAACAGACTCA 4053

Qy    1498 CAATATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAAT 1557
           |||||||   | |||||  ||  |  ||    |   ||     | |||| | |
Db    4054 CAATATGCATTAGGAATCATTCAAGCACAACCAGATAAGAGTGAATCAGAGTTAGTCAGT 4113

Qy    1558 CAAATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACAC 1617
           ||||||||||  |  | || ||||||  | |  | ||  | |||||||||||||||||
Db    4114 CAAATAATAGAGCAGTTAATAAAAAAGGAAAAGGTCTACCTGGCATGGGTACCAGCACAC 4173

Qy    1618 AAAGGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTT 1677
           ||||| || |||||||  || || ||||  |  |||| |||   || || ||  ||||
Db    4174 AAAGGAATTGGAGGAAATGAACAAGTAGATAAATTAGTCAGTGCTGGAATCAGGAAAGTA 4233

Qy    1678 CTCTTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTA 1737
           || || ||| |   |||||   || ||||||||||||| ||||| || ||||||
Db    4234 CTATTTTTGAATGGAATAGATAAGGCCCAAGAAGAACATGAGAAATATCACAGTAATTGG 4293

Qy    1738 AAAGAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACC 1797
           | || | |||      | |||    | ||     |||| || | | | |||||||| || |
Db    4294 AGAGCAATGGCTAGTGATTTTAACCTGCCACCTGTAGTAGCAAAAGAAATAGTAGCCAGC 4353

Qy    1798 TGTGATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGG 1857
           |||||||||||||| |  |||||||| || || ||||| || | | | |    || |||
Db    4354 TGTGATAAATGTCAGCTAAAAGGAGAAGCCATGCATGGACAAGTAGACTGTAGTCCAGGA 4413

Qy    1858 ACTTGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTA 1917
           |  |||||| | |||||||| |||||||| |||||||| ||| | || ||||| ||||||
Db    4414 ATATGGCAACTAGATTGTACACATCTAGAAGGAAAAATTATCCTGGTAGCAGTTCATGTA 4473

Qy    1918 GCTAGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTA 1977
           || |||||||  |||||||||| || ||||||  |||||||||      | ||||||
Db    4474 GCCAGTGGATATATAGAAGCAGAAGTTATTCCAGCAGAGACAGGGCAGGAAACAGCATAT 4533

Qy    1978 TTTCTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCT 2037
           ||||| |||||||| ||||| |||||||||  | | |    |||| ||||| ||||||
Db    4534 TTTCTCTTAAAATTAGCAGGAAGATGGCCAGTAAAAACAATACATACAGACAATGGCAGC 4593

Qy    2038 AACTTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTT 2097
           || ||  |          || |||    |      |||||||||||||  ||||  |  |||
Db    4594 AATTTCACCAGTACTACGGTTAAGGCCGCCTGTTGGTGGGCAGGGATCAAGCAGGAATTT 4653
```

```
Qy   2098 GGGGTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAA 2157
          ||  | || |||||||| || ||||| |||||||| ||| | |||||| |  |  | ||
Db   4654 GGCATTCCCTACAATCCCCAAAGTCAAGGAGTAGTAGAATCTATGAATAATGAATTAAAG 4713

Qy   2158 AATCAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCA 2217
          ||    ||||       | || || || ||  |      |  | ||   ||||  |||||||
Db   4714 AAAATTATAGGACAGGTAAGAGATCAGGCTGAACACCTTAAGACAGCAGTACAAATGGCA 4773

Qy   2218 GTTCATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGA 2277
          ||      |   ||||||||||||||  || || || ||||     | | |||  ||||||
Db   4774 GTATTCATCCACAATTTTAAAAGAAAAGGGGGGATTGGGGGATACAGTGCAGGGGAAAGA 4833

Qy   2278 TTAATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAA 2337
           || |  |||| ||  | |||||   | | |   ||   | ||||    |||   |||||
Db   4834 ATAGTAGACATAATAGCAACAGACATACAAACTAAAGAACTACAAAAGCAAATTACAAAA 4893

Qy   2338 TTTAAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGT 2397
          || |||||||||||||| |||||||| ||   || |||||  || ||||| ||||| |
Db   4894 ATTCAAAATTTTCGGGTTTATTACAGGGACAACAAAGATCCCCTTTGGAAAGGACCAGCA 4953

Qy   2398 GAGCTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTA 2457
          ||||  | |||||||| |||||  |||||  .|  || | |      |  ||||| || |||
Db   4954 AAGCTTCTCTGGAAAGGTGAAGGGGCAGTAGTAATACAAGATAATAGTGACATAAAAGTA 5013

Qy   2458 GTACCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGAT 2514
          || || |||||||| || ||||| || |  |||||||||  |    |  |  |  |||
Db   5014 GTGCCAAGAAGAAAAGCAAAAATCATTAGGGATTATGGAAAACAGATGGCAGGTGAT 5070
```

Search completed: November 16, 2002, 07:57:46
Job time : 275.435 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 16, 2002, 07:33:33 ; Search time 65.1986 Seconds
(without alignments)
14531.422 Million cell updates/sec

Title: M33262_COPY_3105_5668
Perfect score: 2564
Sequence: 1 TTTCCCATAGCTAAAGTAGA.....GCTAGAGAGGTGGCATAGCC 2564

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 328367 seqs, 184756068 residues

Total number of hits satisfying chosen parameters: 656734

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1070.2	41.7	9793	9	US-09-886-156-56	Sequence 56, Appl
2	1069	41.7	9719	10	US-09-737-190A-1	Sequence 1, Appli
3	1067.4	41.6	4307	10	US-09-999-183-1	Sequence 1, Appli
4	1067.4	41.6	8933	10	US-09-943-286-3	Sequence 3, Appli
5	1062.6	41.4	8933	10	US-09-943-286-4	Sequence 4, Appli
6	1056.2	41.2	8933	10	US-09-943-286-9	Sequence 9, Appli
7	961	37.5	2507	10	US-09-872-733-2	Sequence 2, Appli
8	958.4	37.4	2348	9	US-10-097-997-1	Sequence 1, Appli
9	888.2	34.6	8366	10	US-09-872-733-6	Sequence 6, Appli
10	862.8	33.7	4338	10	US-09-872-733-1	Sequence 1, Appli
11	843	32.9	2467	10	US-09-872-733-3	Sequence 3, Appli
12	832.2	32.5	4307	10	US-09-999-183-2	Sequence 2, Appli
13	775.4	30.2	2601	10	US-09-735-487-7	Sequence 7, Appli
14	775.4	30.2	2601	10	US-09-735-487-9	Sequence 9, Appli
15	775.4	30.2	2601	10	US-09-735-487-11	Sequence 11, Appl
16	775.4	30.2	2601	10	US-09-735-487-13	Sequence 13, Appl
17	705.2	27.5	9544	10	US-09-798-675-4	Sequence 4, Appli
18	705.2	27.5	9918	10	US-09-798-675-5	Sequence 5, Appli
19	543.8	21.2	12379	10	US-09-991-258-14	Sequence 14, Appl
20	542.4	21.2	1323	10	US-09-991-258-15	Sequence 15, Appl
21	387.6	15.1	948	10	US-09-756-551A-5	Sequence 5, Appli
22	386.6	15.1	9751	10	US-09-946-239-7	Sequence 7, Appli
23	378.6	14.8	13049	10	US-09-231-235-1	Sequence 1, Appli
24	378.6	14.8	13049	10	US-09-797-518A-1	Sequence 1, Appli
25	354	13.8	9464	10	US-09-738-847-1	Sequence 1, Appli
26	113.2	4.4	4188	10	US-09-920-552-20	Sequence 20, Appl
27	113.2	4.4	4961	10	US-09-920-552-15	Sequence 15, Appl
c 28	113.2	4.4	4961	10	US-09-920-552-18	Sequence 18, Appl
29	110	4.3	1896	10	US-09-920-552-5	Sequence 5, Appli
c 30	110	4.3	1896	10	US-09-920-552-7	Sequence 7, Appli
31	110	4.3	3000	10	US-09-920-552-8	Sequence 8, Appli
c 32	110	4.3	3000	10	US-09-920-552-11	Sequence 11, Appl
33	106.8	4.2	4160	10	US-09-920-552-19	Sequence 19, Appl
34	106.8	4.2	6440	10	US-09-920-552-127	Sequence 127, App
35	100.2	3.9	932	10	US-09-920-552-49	Sequence 49, Appl
c 36	100.2	3.9	932	10	US-09-920-552-52	Sequence 52, Appl
37	83.4	3.3	1985	10	US-09-864-761-4560	Sequence 4560, Ap
38	77.2	3.0	1345	10	US-09-842-883-1	Sequence 1, Appli
39	71.8	2.8	1716	10	US-09-960-428-4	Sequence 4, Appli
40	71.8	2.8	2574	10	US-09-960-428-5	Sequence 5, Appli
41	70.2	2.7	156	10	US-09-860-996-20	Sequence 20, Appl
c 42	65.4	2.6	598	10	US-09-864-761-6993	Sequence 6993, Ap
c 43	64.2	2.5	497	10	US-09-864-761-23724	Sequence 23724, A
44	62.8	2.4	593	10	US-09-864-761-21306	Sequence 21306, A
c 45	52.4	2.0	1643	9	US-10-001-887-41	Sequence 41, Appl

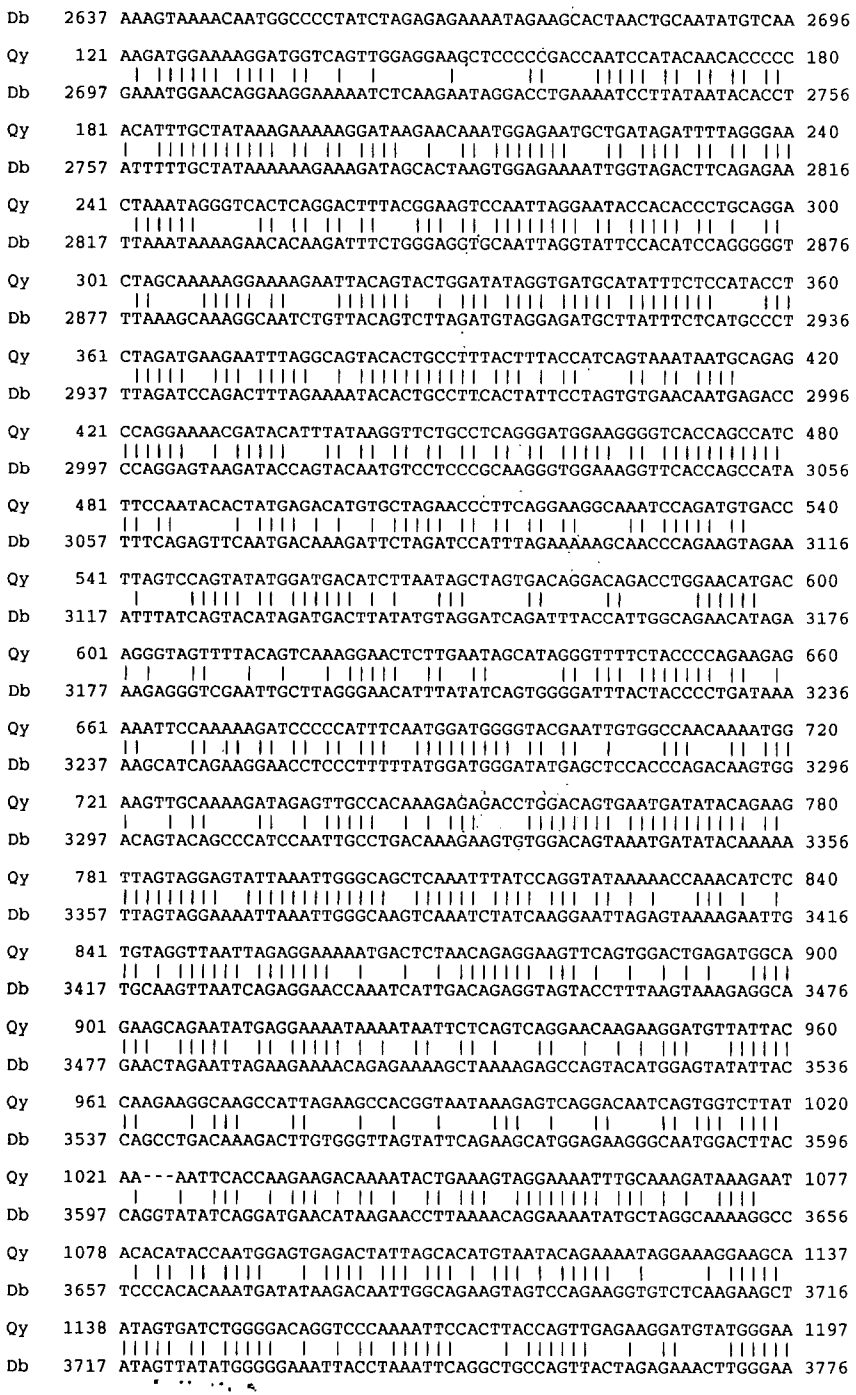
ALIGNMENTS

RESULT 1
US-09-886-156-56
; Sequence 56, Application US/09886156
; Patent No. US20020155428A1
; GENERAL INFORMATION:
; APPLICANT: Guertler, Lutz G.
; APPLICANT: Eberle, Josef
; APPLICANT: Brunn, Albrecht V.
; APPLICANT: Knapp, Stefan
; APPLICANT: Hauser, Hans-Peter
; TITLE OF INVENTION: RETROVIRUS FROM HIV GROUP AND ITS USE
; FILE REFERENCE: 05495.0001-04
; CURRENT APPLICATION NUMBER: US/09/886,156
; CURRENT FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: US/09/109,916
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: DE P 42 33 646.5
; PRIOR FILING DATE: 1992-10-06
; PRIOR APPLICATION NUMBER: DE P 42 35 718.7
; PRIOR FILING DATE: 1992-10-22
; PRIOR APPLICATION NUMBER: DE P 42 44 541.8
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: DE P 43 18 186.4
; PRIOR FILING DATE: 1993-06-01
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 56
; LENGTH: 9793
; TYPE: DNA
; ORGANISM: Human immunodeficiency virus
US-09-886-156-56

Query Match 41.7%; Score 1070.2; DB 9; Length 9793;
Best Local Similarity 64.5%; Pred. No. 1.2e-247;
Matches 1613; Conservative 0; Mismatches 883; Indels 3; Gaps 1;

```
Qy       1 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 60
           || || ||| | ||| || ||  ||| |||| ||||||| ||||||||||
Db    2577 TTCCCTATAAGTCCCATAGCCCCAGTGCCAGTAAAGCTAAAACCAGGAATGGATGGACCA 2636

Qy      61 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 120
           ||| | || || ||||| |||| | ||| || |||| ||| |||  | ||| ||| ||
```


Db 2637 AAAGTAAAACAATGGCCCCTATCTAGAGAGAAAATAGAAGCACTAACTGCAATATGTCAA 2696
Qy 121 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 180
Db 2697 GAAATGGAACAGGAAGGAAAAATCTCAAGAATAGGACCTGAAAATCCTTATAATACACCT 2756
Qy 181 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 240
Db 2757 ATTTTTGCTATAAAAAAGAAAGATAGCACTAAGTGGAGAAAATTGGTAGACTTCAGAGAA 2816
Qy 241 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 300
Db 2817 TTAAATAAAAGAACACAAGATTTCTGGGAGGTGCAATTAGGTATTCCACATCCAGGGGGT 2876
Qy 301 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 360
Db 2877 TTAAAGCAAAGGCAATCTGTTACAGTCTTAGATGTAGGAGATGCTTATTTCTCATGCCCT 2936
Qy 361 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 420
Db 2937 TTAGATCCAGACTTTAGAAAATACACTGCCTTCACTATTCCTAGTGTGAACAATGAGACC 2996
Qy 421 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 480
Db 2997 CCAGGAGTAAGATACCAGTACAATGTCCTCCCGCAAGGGTGGAAAGGTTCACCAGCCATA 3056
Qy 481 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 540
Db 3057 TTTCAGAGTTCAATGACAAAGATTCTAGATCCATTTAGAAAAAGCAACCCAGAAGTAGAA 3116
Qy 541 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 600
Db 3117 ATTTATCAGTACATAGATGACTTATATGTAGGATCAGATTTACCATTGGCAGAACATAGA 3176
Qy 601 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 660
Db 3177 AAGAGGGTCGAATTGCTTAGGGAACATTTATATCAGTGGGGATTTACTACCCCTGATAAA 3236
Qy 661 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 720
Db 3237 AAGCATCAGAAGGAACCTCCCTTTTTATGGATGGGATATGAGCTCCACCCAGACAAGTGG 3296
Qy 721 AAGTTGCAAAAGATAGAGTTGCCACAAAGACAGACCTGGACAGTGAATGATATACAGAAG 780
Db 3297 ACAGTACAGCCCATCCAATTGCCTGACAAAGAAGTGTGGACAGTAAATGATATACAAAAA 3356
Qy 781 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 840
Db 3357 TTAGTAGGAAAATTAAATTGGGCAAGTCAAATCTATCAAGGAATTAGAGTAAAAGAATTG 3416
Qy 841 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 900
Db 3417 TGCAAGTTAATCAGAGGAACCAAATCATTGACAGAGGTAGTACCTTTAAGTAAAGAGGCA 3476
Qy 901 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 960
Db 3477 GAACTAGAATTAGAAGAAAACAGAGAAAAGCTAAAAGAGCCAGTACATGGAGTATATTAC 3536
Qy 961 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 1020
Db 3537 CAGCCTGACAAAGACTTGTGGGTTAGTATTCAGAAGCATGGAGAAGGGCAATGGACTTAC 3596
Qy 1021 AA---AATTCACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAAT 1077
Db 3597 CAGGTATATCAGGATGAACATAAGAACCTTAAAACAGGAAAATATGCTAGGCAAAAGGCC 3656
Qy 1078 ACACATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCA 1137
Db 3657 TCCCACACAAATGATATAAGACAATTGGCAGAAGTAGTCCAGAAGGTGTCTCAAGAAGCT 3716
Qy 1138 ATAGTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAA 1197
Db 3717 ATAGTTATATGGGGGAAATTACCTAAATTCAGGCTGCCAGTTACTAGAGAAACTTGGGAA 3776
Qy 1198 CAGTGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACA 1257
Db 3777 ACTTGGTGGGCAGAATATTGGCAGGCCACCTGGATTCCTGAATGGGAATTTGTCAGCACA 3836
Qy 1258 CCACCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACC 1317
Db 3837 CCCCCATTGATCAAATTATGGTACCAGTTAGAAACAGAACCTATTGTAGGGGCAGAAACC 3896
Qy 1318 TATTATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACA 1377
Db 3897 TTTTATGTAGATGGAGCAGCTAATAGGAATACAAAACTAGGAAAGGCGGGATATGTTACA 3956
Qy 1378 GATAGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTG 1437
Db 3957 GAACAAGGAAAACAGAACATAATAAAGTTAGAAGAGACAACCAATCAAAAGGCTGAATTA 4016
Qy 1438 GAAGCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCA 1497
Db 4017 ATGGCTGTATTAATAGCCTTGCAGGATTCCAAGGAGCAAGTAAACATAGTAACAGACTCA 4076
Qy 1498 CAATATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAAT 1557
Db 4077 CAATATGTATTGGGCATCATATCCTCCCAACCAACACAGAGTGACTCCCCTATAGTTCAG 4136
Qy 1558 CAAATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACAC 1617
Db 4137 CAGATAATAGAGGAACTAACAAAAAAGGAACGAGTGTATCTTACATGGGTTCCTGCTCAC 4196
Qy 1618 AAAGGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTT 1677
Db 4197 AAAGGCATAGGAGGAAATGAAAAAATAGATAAATTAGTAAGCAAAGACATTAGAAGAGTC 4256
Qy 1678 CTCTTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTA 1737
Db 4257 CTGTTCCTGGAAGGAATAGATCAGGCACAAGAAGATCATGAAAAATATCATAGTAATTGG 4316
Qy 1738 AAAGAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACC 1797
Db 4317 AGAGCATTAGCTAGTGACTTTGGATTACCACCAATAGTAGCCAAGGAAATCATTGCTAGT 4376
Qy 1798 TGTGATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGG 1857
Db 4377 TGTCCTAAATGCCATATAAAAGGGGAAGCAACGCATGGTCAAGTAGACTACAGCCCAGAG 4436
Qy 1858 ACTTGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTA 1917
Db 4437 ATATGGCAAATGGATTGTACACATTTAGAAGGCAAAATCATAATAGTTGCTGTCCATGTA 4496
Qy 1918 GCTAGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTA 1977
Db 4497 GCAAGTGACTTTATAGAAGCAGAGGTGATACCAGCAGAAACAGGACAGGAAACTGCCTAT 4556
Qy 1978 TTTCTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCT 2037
Db 4557 TTCCTGTTAAAATTAGCAGCAAGATGGCCTGTCAAAGTAATACATACAGACAATGGACCT 4616
Qy 2038 AACTTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTT 2097
Db 4617 AATTTTACAAGTGCAGCCATGAAAGCTGCATGTTGGTGGACAGGCATACAACATGAGTTT 4676
Qy 2098 GGGGTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAA 2157
Db 4677 GGGATACCATATAATCCACAAAGTCAAGGAGTAGTAGAAGCCATGAATAAAGAATTAAAA 4736
Qy 2158 AATCAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCA 2217
Db 4737 TCTATTATACAGCAGGTGAGGGACCAAGCAGAGCATTTAAAAACAGCAGTACAAATGGCA 4796
Qy 2218 GTTCATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGA 2277
Db 4797 GTCTTTGTTCACAATTTTAAAAGAAAAGGGGGGATTGGGGGGTACACTGCAGGGGAGAGA 4856

```
Qy   2278 TTAATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAA 2337
          |||| |||| |  |  || ||  | | | |  |   | |||| | |||  | ||||
Db   4857 CTAATAGACATACTAGCATCACAAATACAAACAACAGAACTACAAAAACAAATTTTAAAA 4916

Qy   2338 TTTAAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGT 2397
          | || ||||||||||||||||||||||||  ||||||| |   | ||||| ||||| |
Db   4917 ATCAACAATTTTCGGGTCTATTACAGAGATAGCAGAGACCCTATTTGGAAAGGACCGGCA 4976

Qy   2398 GAGCTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTA 2457
          | ||  |||||||||| || || |||||  || || | |       ||||||||| ||
Db   4977 CAACTCCTGTGGAAAGGTGAGGGGGCAGTAGTCATACAAGATAAAGGAGACATTAAAGTG 5036

Qy   2458 GTACCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGA 2496
          ||||| |||||||||| ||||| |||| ||||||||||
Db   5037 GTACCAAGAAGAAAGGCAAAAATAATCAGAGATTATGGA 5075


RESULT 2
US-09-737-190A-1
; Sequence 1, Application US/09737190A
; Patent No. US20020102545A1
; GENERAL INFORMATION:
;  APPLICANT: Shibuya, Tetsuo
;  TITLE OF INVENTION: A Method for Changing a Target Array, a Method for Analyzing
;  TITLE OF INVENTION:  a Structure, and an Apparatus, a Storage Medium and a
;  TITLE OF INVENTION:  Transmission Medium Therefor
;  FILE REFERENCE: JP919990270US1 (14043)
;  CURRENT APPLICATION NUMBER: US/09/737,190A
;  CURRENT FILING DATE:  2000-12-14
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 9719
;   TYPE: DNA
;   ORGANISM: Human iImmunodeficiency virus type 1
US-09-737-190A-1

  Query Match             41.7%;  Score 1069;  DB 10;  Length 9719;
  Best Local Similarity   64.3%;  Pred. No. 2.4e-247;
  Matches 1619;  Conservative    0;  Mismatches  895;  Indels    3;  Gaps    1;

Qy      1 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 60
          ||||||||        | ||| |||||  |||    ||||||||||||| ||||||  |||
Db   2547 TTTCCCATTAGCCCTATTGAGACTGTACCAGTAAAATTAAAGCCAGGAATGGATGGCCCA 2606

Qy     61 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 120
          ||| | || || |||||||  || ||||||| |||   ||||||  |||||| |||  |
Db   2607 AAAGTTAAACAATGGCCATTGACAGAAGAAAAAATAAAAGCATTAGTAGAAATTTGTACA 2666

Qy    121 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 180
          ||||||||||||| ||  |  |    ||  |  ||   |||||||||| || ||
Db   2667 GAGATGGAAAAGGAAGGGAAAATTTCAAAAATTGGGCCTGAAAATCCATACAATACTCCA 2726

Qy    181 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 240
            |||||| |||||||||| || |  |  |||||||||   |  |||||| || |||
Db   2727 GTATTTGCCATAAAGAAAAAAGACAGTACTAAATGGAGAAAATTAGTAGATTTCAGAGAA 2786

Qy    241 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 300
          || |||| |   ||||| |||||   |||||| |||||||||||||||| || |||||
Db   2787 CTTAATAAGAGAACTCAAGACTTCTGGGAAGTTCAATTAGGAATACCACATCCCGCAGGG 2846

Qy    301 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 360
          ||  ||| |  |||  | | |||||||||||| | ||||||||||||| ||  | ||
Db   2847 TTAAAAAAGAAAAAATCAGTAACAGTACTGGATGTGGGTGATGCATATTTTTCAGTTCCC 2906

Qy    361 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 420
          |||||||||| || ||| |||| ||||| |||||  ||||     |||| ||||
Db   2907 TTAGATGAAGACTTCAGGAAGTATACTGCATTTACCATACCTAGTATAAACAATGAGACA 2966

Qy    421 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 480
          ||||| |   ||||    || || || || || |||||||||| || ||||||||  ||
Db   2967 CCAGGGATTAGATATCAGTACAATGTGCTTCCACAGGGATGGAAAGGATCACCAGCAATA 3026

Qy    481 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 540
          ||||||   |  |||| | |  |  |||| || || || ||  |||||||||  ||
Db   3027 TTCCAAAGTAGCATGACAAAAATCTTAGAGCCTTTTAGAAAACAAAATCCAGACATAGTT 3086

Qy    541 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 600
          |    || || ||||||||  | |   |||   ||||    |   | | |||
Db   3087 ATCTATCAATACATGGATGATTTGTATGTAGGATCTGACTTAGAAATAGGGCAGCATAGA 3146

Qy    601 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 660
          |    | |     ||   | |||   ||   ||    ||  || | || |||||  |
Db   3147 ACAAAAATAGAGGAGCTGAGACAACATCTGTTGAGGTGGGGACTTACCACACCAGACAAA 3206

Qy    661 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 720
          |||   || ||||| || ||||| |  |||||||| || ||| |    ||    ||||||
Db   3207 AAACATCAGAAAGAACCTCCATTCCTTTGGATGGGTTATGAACTCCATCCTGATAAATGG 3266

Qy    721 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 780
          |   | ||    |||| | ||||| ||| ||| | |||||| || |||||  ||||||||
Db   3267 ACAGTACAGCCTATAGTGCTGCCAGAAAAAGACAGCTGGACTGTCAATGACATACAGAAG 3326

Qy    781 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 840
          ||||| ||   ||| |||||||||  ||| ||||| ||||| || |||   |  ||  |
Db   3327 TTAGTGGGGAAATTGAATTGGGCAAGTCAGATTTACCCAGGGATTAAAGTAAGGCAATTA 3386

Qy    841 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 900
          ||||   | |||||||||   | ||||||| | | |    || ||  | |||
Db   3387 TGTAAACTCCTTAGAGGAACCAAAGCACTAACAGAAGTAATACCACTAACAGAAGAAGCA 3446

Qy    901 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 960
          ||   ||||   | ||||| | |   ||||| |   |   | | | |||   ||||||
Db   3447 GAGCTAGAACTGGCAGAAAACAGAGAGATTCTAAAAGAACCAGTACATGGAGTGTATTAT 3506

Qy    961 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 1020
          |   |  ||   ||| ||| .  ||  |||     | |    || ||| || |||
Db   3507 GACCCATCAAAAGACTTAATAGCAGAAATACAGAAGCAGGGGCAAGGCCAATGGACATAT 3566

Qy   1021 AAAATTCACCAAGA---AGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAAT 1077
          ||||| | |||||   |  ||||  ||||||  |||||||| ||||  || | | |
Db   3567 CAAATTTATCAAGAGCCATTTAAAAATCTGAAAACAGGAAAATATGCAAGAATGAGGGGT 3626

Qy   1078 ACACATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCA 1137
          | || || |||| || | || |||| || | | | || ||||||   |  |||
Db   3627 GCCCACACTAATGATGTAAAACAATTAACAGAGGCAGTGCAAAAAATAACCACAGAAAGC 3686

Qy   1138 ATAGTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAA 1197
          ||||| || |||||| ||   || |||||  | | ||  |  | |||||   |||||||
Db   3687 ATAGTAATATGGGGAAAGACTCCTAAATTTAAACTGCCCATACAAAAGGAAACATGGGAA 3746

Qy   1198 CAGTGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACA 1257
             |||||||||| |||||||| |  |||||||| || || ||||| ||| |   ||
Db   3747 ACATGGTGGACAGAGTATTGGCAAGCCACCTGGATTCCTGAGTGGGAGTTTGTTAATACC 3806

Qy   1258 CCACCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACC 1317
          || ||  |||| | ||||   | | |  ||| ||| || || ||||  |||| ||||||
Db   3807 CCTCCCTTAGTGAAATTATGGTACCAGTTAGAGAAAGAACCCATAGTAGGAGCAGAAACC 3866

Qy   1318 TATTATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACA 1377
          |  |||  |||||  ||  ||| |   || | |||  ||| ||||||||||||| | ||
Db   3867 TTCTATGTAGATGGGGCAGCTAACAGGGAGACTAAATTAGGAAAAGCAGGATATGTTACT 3926

Qy   1378 GATAGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTG 1437
          ||||| || | | | |||||        ||    | || || |||||  |  | || ||
Db   3927 AATAGAGGAAGACAAAAAGTTGTCACCCTAACTGACACAACAAATCAGAAGACTGAGTTA 3986

Qy   1438 GAAGCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCA 1497
          |||||  ||    | || |||    || || ||   | | | ||| ||  ||  ||| |||
Db   3987 CAAGCAATTTATCTAGCTTTGCAGGATTCGGGATTAGAAGTAAACATAGTAACAGACTCA 4046

Qy   1498 CAATATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAAT 1557
          |||||||   | |||||  ||  || |   ||    ||    ||       | |||| |||
Db   4047 CAATATGCATTAGGAATCATTCAAGCACAACCAGATCAAAGTGAATCAGAGTTAGTCAAT 4106

Qy   1558 CAAATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACAC 1617
```

```
Db   4107 CAAATAATAGAGCAGTTAATAAAAAAGGAAAAGGTCTATCTGGCATGGGTACCAGCACAC 4166

Qy   1618 AAAGGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTT 1677

Db   4167 AAAGGAATTGGAGGAAATGAACAAGTAGATAAATTAGTCAGTGCTGGAATCAGGAAAGTA 4226

Qy   1678 CTCTTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTA 1737

Db   4227 CTATTTTTAGATGGAATAGATAAGGCCCAAGATGAACATGAGAAATATCACAGTAATTGG 4286

Qy   1738 AAAGAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACC 1797

Db   4287 AGAGCAATGGCTAGTGATTTTAACCTGCCACCTGTAGTAGCAAAAGAAATAGTAGCCAGC 4346

Qy   1798 TGTGATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGG 1857

Db   4347 TGTGATAAATGTCAGCTAAAAGGAGAAGCCATGCATGGACAAGTAGACTGTAGTCCAGGA 4406

Qy   1858 ACTTGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTA 1917

Db   4407 ATATGGCAACTAGATTGTACACATTTAGAAGGAAAAGTTATCCTGGTAGCAGTTCATGTA 4466

Qy   1918 GCTAGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTA 1977

Db   4467 GCCAGTGGATATATAGAAGCAGAAGTTATTCCAGCAGAAACAGGGCAGGAAACAGCATAT 4526

Qy   1978 TTTCTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCT 2037

Db   4527 TTTCTTTTAAAATTAGCAGGAAGATGGCCAGTAAAAACAATACATACTGACAATGGCAGC 4586

Qy   2038 AACTTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTT 2097

Db   4587 AATTTCACCGGTGCTACGGTTAGGGCCGCCTGTTGGTGGGCGGGAATCAAGCAGGAATTT 4646

Qy   2098 GGGGTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAA 2157

Db   4647 GGAATTCCCTACAATCCCCAAAGTCAAGGAGTAGTAGAATCTATGAATAAAGAATTAAAG 4706

Qy   2158 AATCAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCA 2217

Db   4707 AAAATTATAGGACAGGTAAGAGATCAGGCTGAACATCTTAAGACAGCAGTACAAATGGCA 4766

Qy   2218 GTTCATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGA 2277

Db   4767 GTATTCATCCACAATTTTAAAAGAAAAGGGGGGATTGGGGGGTACAGTGCAGGGGAAAGA 4826

Qy   2278 TTAATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAA 2337

Db   4827 ATAGTAGACATAATAGCAACAGACATACAAACTAAAGAATTACAAAAACAAATTACAAAA 4886

Qy   2338 TTTAAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGT 2397

Db   4887 ATTCAAAATTTTCGGGTTTATTACAGGGACAGCAGAAATCCACTTTGGAAAGGACCAGCA 4946

Qy   2398 GAGCTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTA 2457

Db   4947 AAGCTCCTCTGGAAAGGTGAAGGGGCAGTAGTAATACAAGATAATAGTGACATAAAAGTA 5006

Qy   2458 GTACCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGAT 2514

Db   5007 GTGCCAAGAAGAAAAGCAAAGATCATTAGGGATTATGGAAAACAGATGGCAGGTGAT 5063


RESULT 3
US-09-999-183-1
; Sequence 1, Application US/09999183
; Patent No. US20020147169A1
; GENERAL INFORMATION:
;  APPLICANT: MITROPHANOUS, et al
;  TITLE OF INVENTION: In Vivo Selection Method
;  FILE REFERENCE: 674523-2009
;  CURRENT APPLICATION NUMBER: US/09/999,183
;  CURRENT FILING DATE:  2001-11-29
;  PRIOR APPLICATION NUMBER: PCT/GB00/02136
;  PRIOR FILING DATE: 2000-06-02
;  PRIOR APPLICATION NUMBER: 9912965.2
;  PRIOR FILING DATE: 1999-06-03
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: SeqWin99
; SEQ ID NO 1
;   LENGTH: 4307
;   TYPE: DNA
;   ORGANISM: Human immunodeficiency virus type I
US-09-999-183-1

  Query Match             41.6%;  Score 1067.4;  DB 10;  Length 4307;
  Best Local Similarity   64.3%;  Pred. No. 4.1e-247;
  Matches 1618;  Conservative    0;  Mismatches  896;  Indels    3;  Gaps    1;

Qy      1 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 60

Db   1758 TTTCCCATTAGCCCTATTGAGACTGTACCAGTAAAATTAAAGCCAGGAATGGATGGCCCA 1817

Qy     61 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 120

Db   1818 AAAGTTAAACAATGGCCATTGACAGAAGAAAAAATAAAAGCATTAGTAGAAATTTGTACA 1877

Qy    121 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 180

Db   1878 GAGATGGAAAAGGAAGGGAAAATTTCAAAAATTGGGCCTGAAAATCCATACAATACTCCA 1937

Qy    181 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 240

Db   1938 GTATTTGCCATAAAGAAAAAAGACAGTACTAAATGGAGAAAATTAGTAGATTTCAGAGAA 1997

Qy    241 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 300

Db   1998 CTTAATAAGAGAACTCAAGACTTCTGGGAAGTTCAATTAGGAATACCACATCCCGCAGGG 2057

Qy    301 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 360

Db   2058 TTAAAAAAGAAAAAATCAGTAACAGTACTGGATGTGGGTGATGCATATTTTTCAGTTCCC 2117

Qy    361 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 420

Db   2118 TTAGATGAAGACTTCAGGAAGTATACTGCATTTACCATACCTAGTATAAACAATGAGACA 2177

Qy    421 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 480

Db   2178 CCAGGGATTAGATATCAGTACAATGTGCTTCCACAGGGATGGAAAGGATCACCAGCAATA 2237

Qy    481 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 540

Db   2238 TTCCAAAGTAGCATGACAAAAATCTTAGAGCCTTTTAGAAAACAAAATCCAGACATAGTT 2297

Qy    541 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 600

Db   2298 ATCTATCAATACATGGATGATTTGTATGTAGGATCTGACTTAGAAATAGGGCAGCATAGA 2357

Qy    601 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 660

Db   2358 ACAAAAATAGAGGAGCTGAGACAACATCTGTTGAGGTGGGGACTTACCACACCAGACAAA 2417

Qy    661 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 720

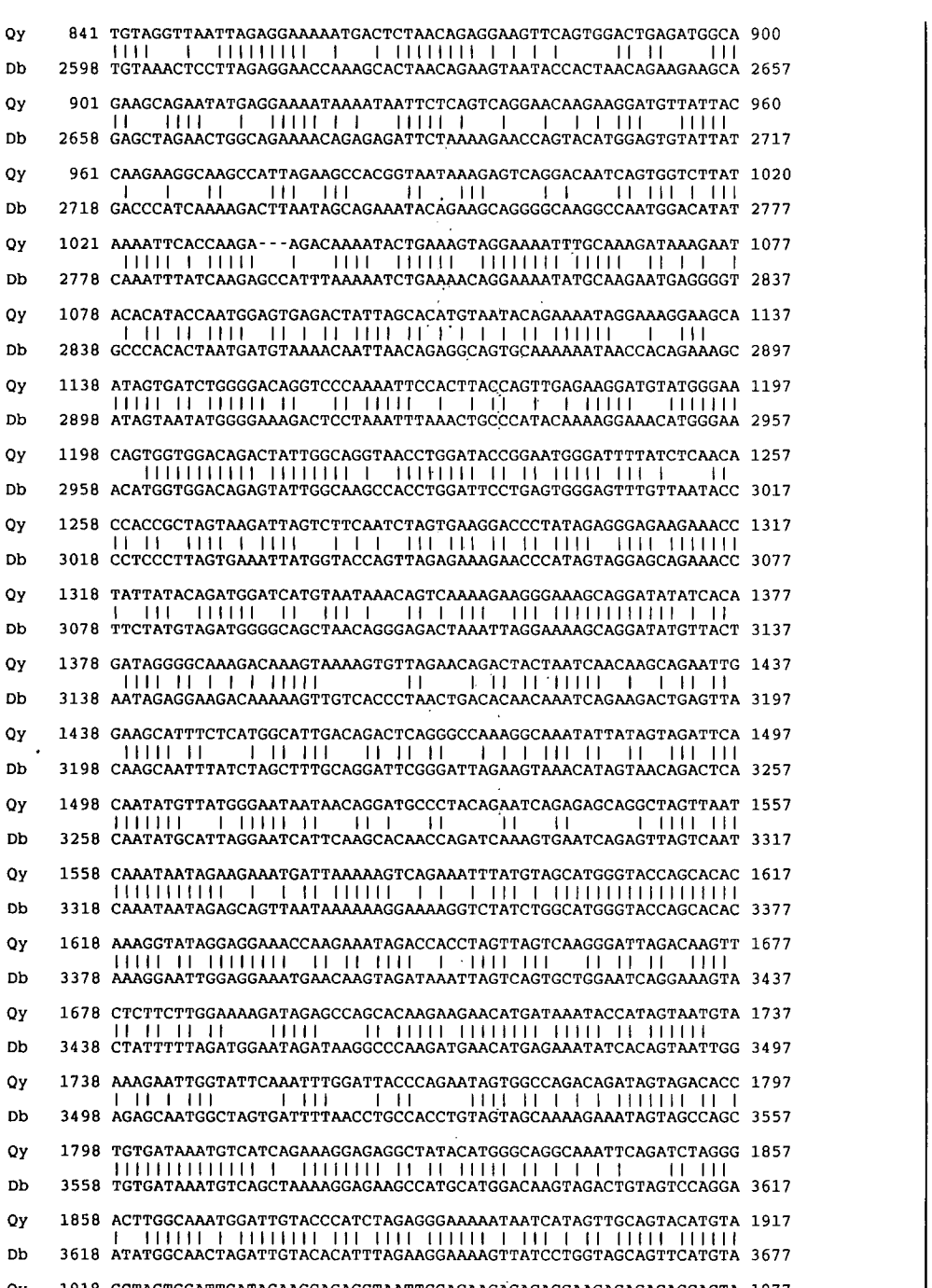
Db   2418 AAACATCAGAAAGAACCTCCATTCCTTTGGATGGGTTATGAACTCCATCCTGATAAATGG 2477

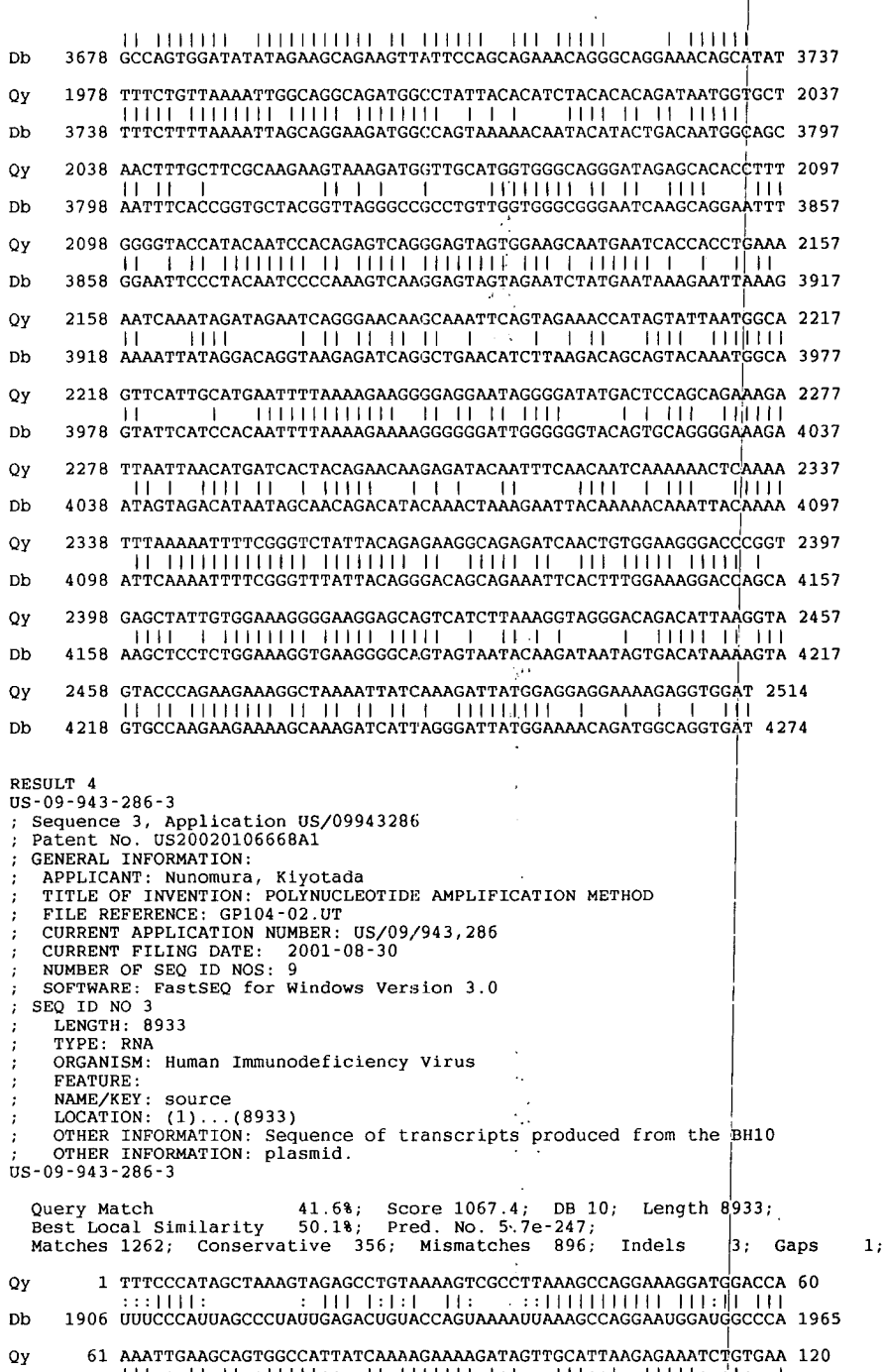
Qy    721 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 780

Db   2478 ACAGTACAGCCTATAGTGCTGCCAGAAAAAGACAGCTGGACTGTCAATGACATACAGAAG 2537

Qy    781 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 840

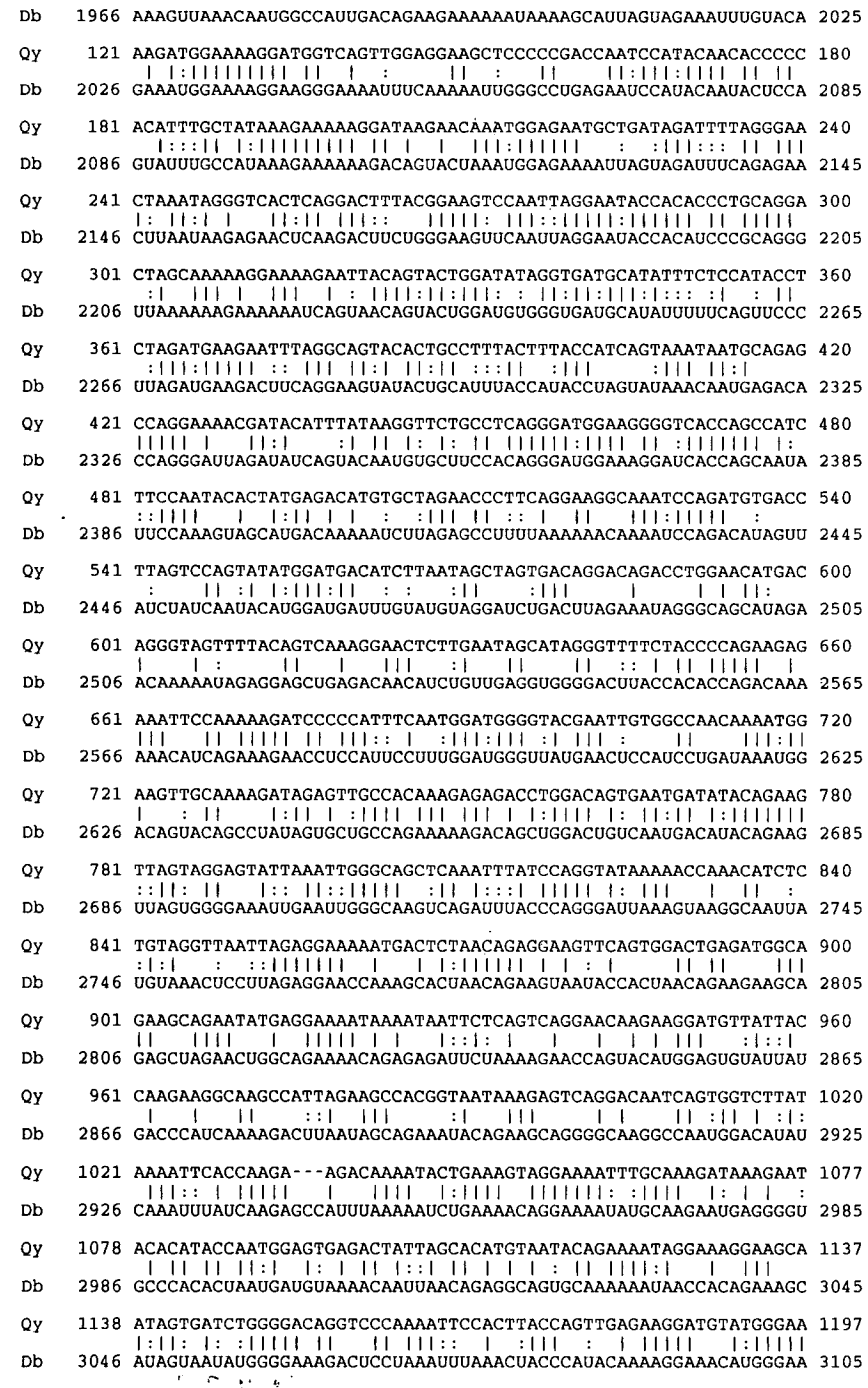
Db   2538 TTAGTGGGGAAATTGAATTGGGCAAGTCAGATTTACCCAGGGATTAAAGTAAGGCAATTA 2597
```

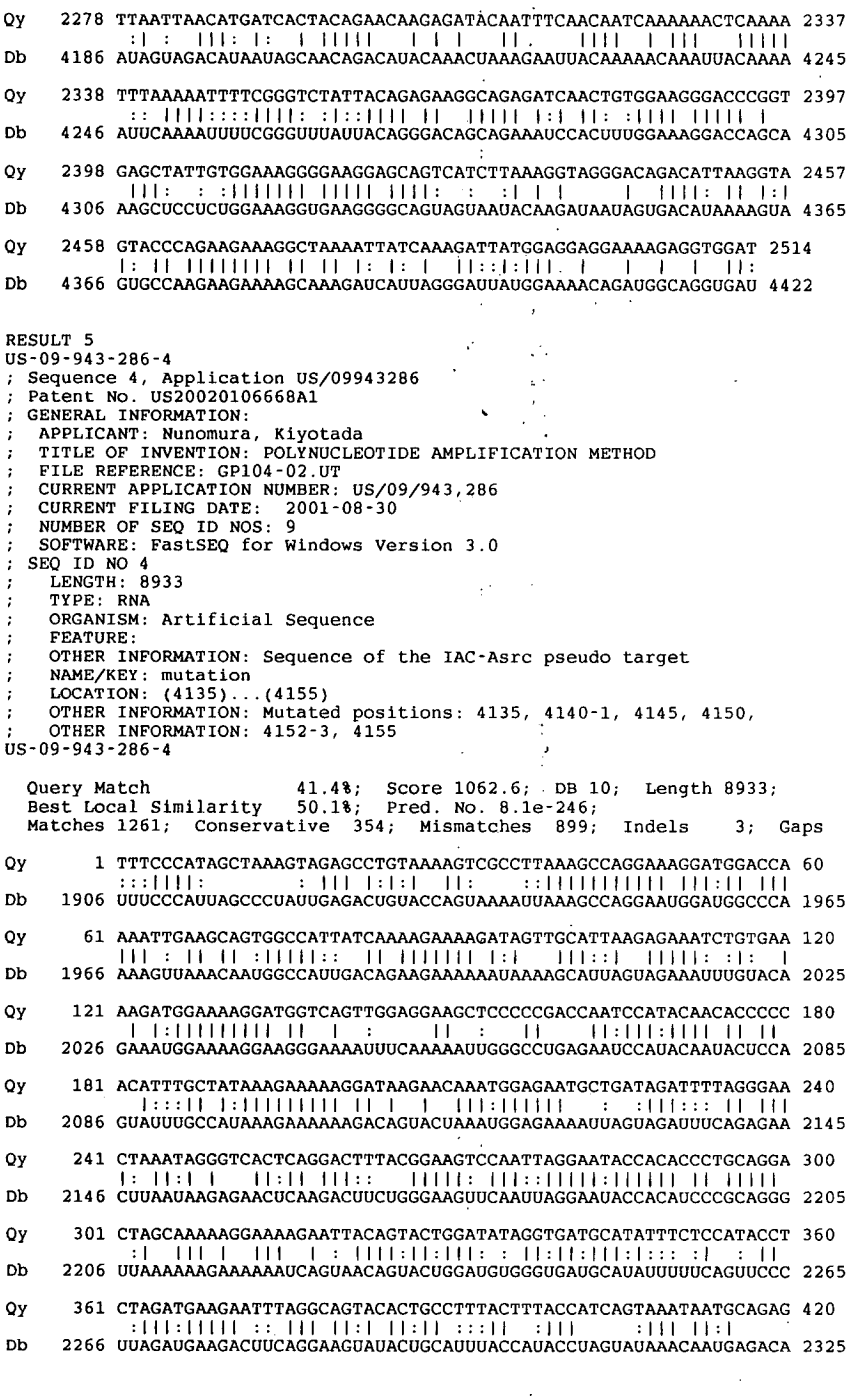




RESULT 4
US-09-943-286-3
; Sequence 3, Application US/09943286
; Patent No. US20020106668A1
; GENERAL INFORMATION:
; APPLICANT: Nunomura, Kiyotada
; TITLE OF INVENTION: POLYNUCLEOTIDE AMPLIFICATION METHOD
; FILE REFERENCE: GP104-02.UT
; CURRENT APPLICATION NUMBER: US/09/943,286
; CURRENT FILING DATE: 2001-08-30
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 8933
; TYPE: RNA
; ORGANISM: Human Immunodeficiency Virus
; FEATURE:
; NAME/KEY: source
; LOCATION: (1)...(8933)
; OTHER INFORMATION: Sequence of transcripts produced from the BH10
; OTHER INFORMATION: plasmid.
US-09-943-286-3

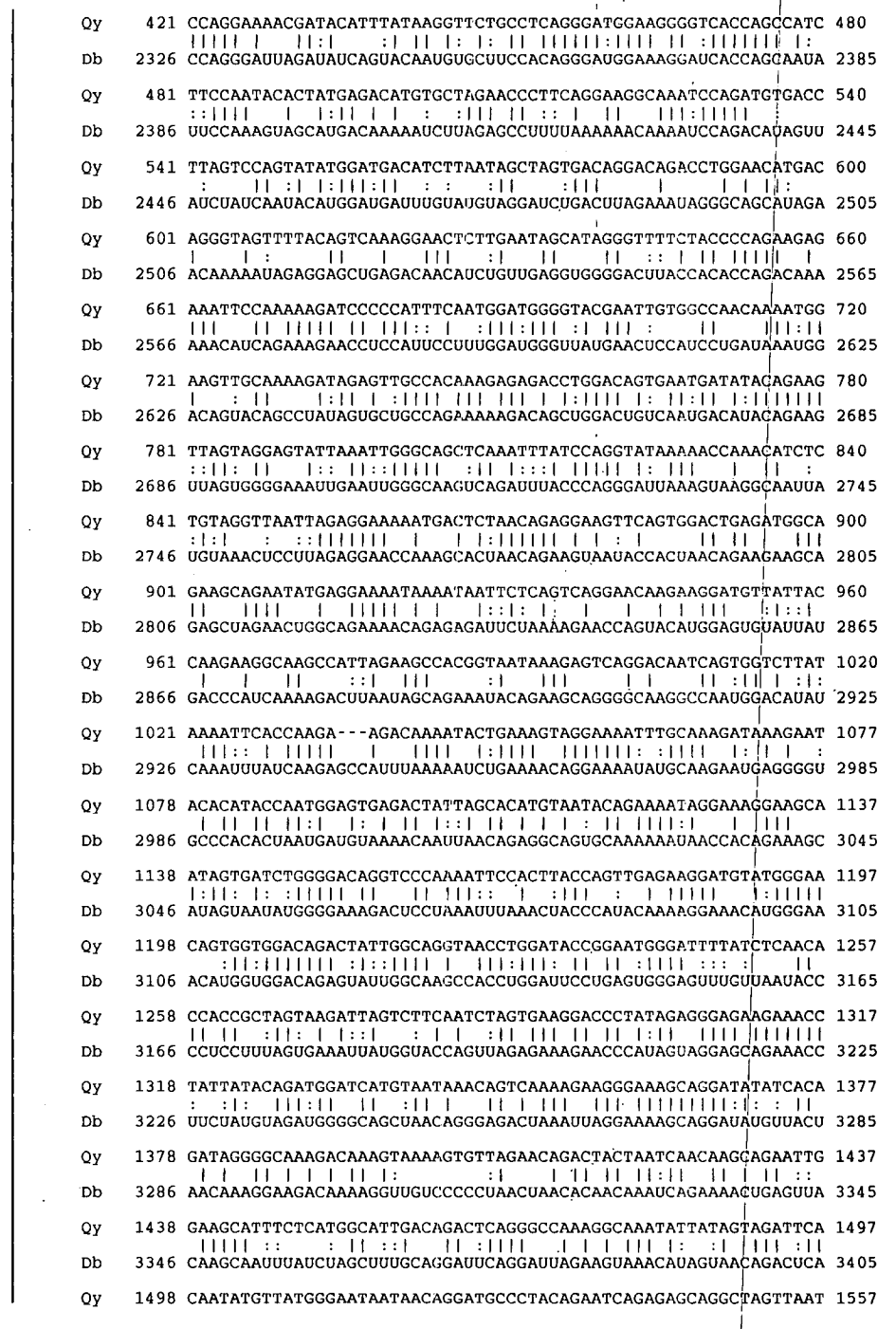
Query Match 41.6%; Score 1067.4; DB 10; Length 8933;
Best Local Similarity 50.1%; Pred. No. 5.7e-247;
Matches 1262; Conservative 356; Mismatches 896; Indels 3; Gaps 1;


Db 1966 AAAGUUAAACAAUGGCCAUUGACAGAAGAAAAAAUAAAAGCAUUAGUAGAAAUUUGUACA 2025
Qy 121 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 180
Db 2026 GAAAUGGAAAAGGAAGGGAAAAUUUCAAAAAUUGGGCCUGAGAAUCCAUACAAUACUCCA 2085
Qy 181 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 240
Db 2086 GUAUUUGCCAUAAAGAAAAAAGACAGUACUAAAUGGAGAAAAUUAGUAGAUUUCAGAGAA 2145
Qy 241 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 300
Db 2146 CUUAAUAAGAGAACUCAAGACUUCUGGGAAGUUCAAUUAGGAAUACCACAUCCCGCAGGG 2205
Qy 301 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 360
Db 2206 UUAAAAAAGAAAAAAUCAGUAACAGUACUGGAUGUGGGUGAUGCAUAUUUUUCAGUUCCC 2265
Qy 361 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 420
Db 2266 UUAGAUGAAGACUUCAGGAAGUAUACUGCAUUUACCAUACCUAGUAUAAACAAUGAGACA 2325
Qy 421 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 480
Db 2326 CCAGGGAUUAGAUAUCAGUACAAUGUGCUUCCACAGGGAUGGAAAGGAUCACCAGCAAUA 2385
Qy 481 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 540
Db 2386 UUCCAAAGUAGCAUGACAAAAAUCUUAGAGCCUUUUAAAAAACAAAAUCCAGACAUAGUU 2445
Qy 541 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 600
Db 2446 AUCUAUCAAUACAUGGAUGAUUUGUAUGUAGGAUCUGACUUAGAAAUAGGGCAGCAUAGA 2505
Qy 601 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 660
Db 2506 ACAAAAAUAGAGGAGCUGAGACAACAUCUGUUGAGGUGGGGACUUACCACACCAGACAAA 2565
Qy 661 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 720
Db 2566 AAACAUCAGAAAGAACCUCCAUUCCUUUGGAUGGGUUAUGAACUCCAUCCUGAUAAAUGG 2625
Qy 721 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 780
Db 2626 ACAGUACAGCCUAUAGUGCUGCCAGAAAAAGACAGCUGGACUGUCAAUGACAUACAGAAG 2685
Qy 781 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 840
Db 2686 UUAGUGGGGAAAUUGAAUUGGGCAAGUCAGAUUUACCCAGGGAUUAAAGUAAGGCAAUUA 2745
Qy 841 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 900
Db 2746 UGUAAACUCCUUAGAGGAACCAAAGCACUAACAGAAGUAAUACCACUAACAGAAGAAGCA 2805
Qy 901 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 960
Db 2806 GAGCUAGAACUGGCAGAAAACAGAGAGAUUCUAAAAGAACCAGUACAUGGAGUGUAUUAU 2865
Qy 961 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 1020
Db 2866 GACCCAUCAAAAGACUUAAUAGCAGAAAUACAGAAGCAGGGGCAAGGCCAAUGGACAUAU 2925
Qy 1021 AAAATTCACCAAGA---AGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAAT 1077
Db 2926 CAAAUUUAUCAAGAGCCAUUUAAAAAUCUGAAAACAGGAAAAUAUGCAAGAAUGAGGGGU 2985
Qy 1078 ACACATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCA 1137
Db 2986 GCCCACACUAAUGAUGUAAAACAAUUAACAGAGGCAGUGCAAAAAAUAACCACAGAAAGC 3045
Qy 1138 ATAGTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAA 1197
Db 3046 AUAGUAAUAUGGGGAAAGACUCCUAAAUUUAAACUACCCAUACAAAAGGAAACAUGGGAA 3105
Qy 1198 CAGTGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACA 1257
Db 3106 ACAUGGUGGACAGAGUAUUGGCAAGCCACCUGGAUUCCUGAGUGGGAGUUUGUUAAUACC 3165
Qy 1258 CCACCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACC 1317
Db 3166 CCUCCUUUAGUGAAAUUAUGGUACCAGUUAGAGAAAGAACCCAUAGUAGGAGCAGAAACC 3225
Qy 1318 TATTATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACA 1377
Db 3226 UUCUAUGUAGAUGGGGCAGCUAACAGGGAGACUAAAUUAGGAAAAGCAGGAUAUGUUACU 3285
Qy 1378 GATAGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTG 1437
Db 3286 AACAAAGGAAGACAAAAGGUUGUCCCCCUAACUAACACAACAAAUCAGAAAACUGAGUUA 3345
Qy 1438 GAAGCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCA 1497
Db 3346 CAAGCAAUUUAUCUAGCUUUGCAGGAUUCAGGAUUAGAAGUAAACAUAGUAACAGACUCA 3405
Qy 1498 CAATATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAAT 1557
Db 3406 CAAUAUGCAUUAGGAAUCAUUCAAGCACAACCAGAUAAAAGUGAAUCAGAGUUAGUCAAU 3465
Qy 1558 CAAATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACAC 1617
Db 3466 CAAAUAAUAGAGCAGUUAAUAAAAAAGGAAAAGGUCUAUCUGGCAUGGGUACCAGCACAC 3525
Qy 1618 AAAGGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTT 1677
Db 3526 AAAGGAAUUGGAGGAAAUGAACAAGUAGAUAAAUUAGUCAGUGCUGGAAUCAGGAAAAUA 3585
Qy 1678 CTCTTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTA 1737
Db 3586 CUAUUUUUAGAUGGAAUAGAUAAGGCCCAAGAUGAACAUGAGAAAUAUCACAGUAAUUGG 3645
Qy 1738 AAAGAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACC 1797
Db 3646 AGAGCAAUGGCUAGUGAUUUUAACCUGCCACCUGUAGUAGCAAAAGAAAUAGUAGCCAGC 3705
Qy 1798 TGTGATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGG 1857
Db 3706 UGUGAUAAAUGUCAGCUAAAAGGAGAAGCCAUGCAUGGACAAGUAGACUGUAGUCCAGGA 3765
Qy 1858 ACTTGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTA 1917
Db 3766 AUAUGGCAACUAGAUUGUACACAUUUAGAAGGAAAAGUUAUCCUGGUAGCAGUUCAUGUA 3825
Qy 1918 GCTAGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTA 1977
Db 3826 GCCAGUGGAUAUAUAGAAGCAGAAGUUAUUCCAGCAGAAACAGGGCAGGAAACAGCAUAU 3885
Qy 1978 TTTCTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCT 2037
Db 3886 UUUCUUUUAAAAUUAGCAGGAAGAUGGCCAGUAAAAACAAUACAUACAGACAAUGGCAGC 3945
Qy 2038 AACTTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTT 2097
Db 3946 AAUUUCACCAGUGCUACGGUUAAGGCCGCCUGUUGGUGGGCGGGAAUCAAGCAGGAAUUU 4005
Qy 2098 GGGGTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAA 2157
Db 4006 GGAAUUCCCUACAAUCCCCAAAGUCAAGGAGUAGUAGAAUCUAUGAAUAAAGAAUUAAAG 4065
Qy 2158 AATCAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCA 2217
Db 4066 AAAAUUAUAGGACAGGUAAGAGAUCAGGCUGAACAUCUUAAGACAGCAGUACAAAUGGCA 4125
Qy 2218 GTTCATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGA 2277
Db 4126 GUAUUCAUCCACAAUUUUAAAAGAAAAGGGGGGAUUGGGGGGUACAGUGCAGGGGAAAGA 4185



RESULT 5
US-09-943-286-4
; Sequence 4, Application US/09943286
; Patent No. US20020106668A1
; GENERAL INFORMATION:
; APPLICANT: Nunomura, Kiyotada
; TITLE OF INVENTION: POLYNUCLEOTIDE AMPLIFICATION METHOD
; FILE REFERENCE: GP104-02.UT
; CURRENT APPLICATION NUMBER: US/09/943,286
; CURRENT FILING DATE: 2001-08-30
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 8933
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Sequence of the IAC-Asrc pseudo target
; NAME/KEY: mutation
; LOCATION: (4135)...(4155)
; OTHER INFORMATION: Mutated positions: 4135, 4140-1, 4145, 4150,
; OTHER INFORMATION: 4152-3, 4155
US-09-943-286-4

Query Match 41.4%; Score 1062.6; DB 10; Length 8933;
Best Local Similarity 50.1%; Pred. No. 8.1e-246;
Matches 1261; Conservative 354; Mismatches 899; Indels 3; Gaps 1;



```
Db    3406 CAAUAUGCAUUAGGAAUCAUUCAAGCACAACCAGAUAAAAGUGAAUCAGAGUUAGUCAAU 3465

Qy    1558 CAAATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACAC 1617
Db    3466 CAAAUAAUAGAGCAGUUAAUAAAAAAGGAAAAGGUCUAUCUGGCAUGGGUACCAGCACAC 3525

Qy    1618 AAAGGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTT 1677
Db    3526 AAAGGAAUUGGAGGAAAUGAACAAGUAGAUAAAUUAGUCAGUGCUGGAAUCAGGAAAAUA 3585

Qy    1678 CTCTTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTA 1737
Db    3586 CUAUUUUUAGAUGGAAUAGAUAAGGCCCAAGAUGAACAUGAGAAAUAUCACAGUAAUUGG 3645

Qy    1738 AAAGAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACC 1797
Db    3646 AGAGCAAUGGCUAGUGAUUUUAACCUGCCACCUGUAGUAGCAAAAGAAAUAGUAGCCAGC 3705

Qy    1798 TGTGATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGG 1857
Db    3706 UGUGAUAAAUGUCAGCUAAAAGGAGAAGCCAUGCAUGGACAAGUAGACUGUAGUCCAGGA 3765

Qy    1858 ACTTGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTA 1917
Db    3766 AUAUGGCAACUAGAUUGUACACAUUUAGAAGGAAAAGUUAUCCUGGUAGCAGUUCAUGUA 3825

Qy    1918 GCTAGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTA 1977
Db    3826 GCCAGUGGAUAUAUAGAAGCAGAAGUUAUUCCAGCAGAAACAGGGCAGGAAACAGCAUAU 3885

Qy    1978 TTTCTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCT 2037
Db    3886 UUUCUUUUAAAAUUAGCAGGAAGAUGGCCAGUAAAAACAAUACAUACAGACAAUGGCAGC 3945

Qy    2038 AACTTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTT 2097
Db    3946 AAUUUCACCAGUGCUACGGUUAAGGCCGCCUGUUGGUGGGCGGGAAUCAAGCAGGAAUUU 4005

Qy    2098 GGGGTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAA 2157
Db    4006 GGAAUUCCCUACAAUCCCCAAAGUCAAGGAGUAGUAGAAUCUAUGAAUAAAGAAUUAAAG 4065

Qy    2158 AATCAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCA 2217
Db    4066 AAAAUUAUAGGACAGGUAAGAGAUCAGGCUGAACAUCUUAAGACAGCAGUACAAAUGGCA 4125

Qy    2218 GTTCATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGA 2277
Db    4126 GUAUUCAUCUACAAGCUUAGAAGAUAGAGAGGGAUUGGGGGGUACAGUGCAGGGGAAAGA 4185

Qy    2278 TTAATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAA 2337
Db    4186 AUAGUAGACAUAAUAGCAACAGACAUACAAACUAAAGAAUUACAAAAACAAAUUACAAAA 4245

Qy    2338 TTTAAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGT 2397
Db    4246 AUUCAAAAUUUUCGGGUUUAUUACAGGGACAGCAGAAAUCCACUUUGGAAAGGACCAGCA 4305

Qy    2398 GAGCTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTA 2457
Db    4306 AAGCUCCUCUGGAAAGGUGAAGGGGCAGUAGUAAUACAAGAUAAUAGUGACAUAAAAGUA 4365

Qy    2458 GTACCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGAT 2514
Db    4366 GUGCCAAGAAGAAAAGCAAAGAUCAUUAGGGAUUAUGGAAAACAGAUGGCAGGUGAU 4422

RESULT 6
US-09-943-286-9
; Sequence 9, Application US/09943286
; Patent No. US20020106668A1
; GENERAL INFORMATION:
;  APPLICANT: Nunomura, Kiyotada
;  TITLE OF INVENTION: POLYNUCLEOTIDE AMPLIFICATION METHOD
;  FILE REFERENCE: GP104-02.UT
;  CURRENT APPLICATION NUMBER: US/09/943,286
;  CURRENT FILING DATE:  2001-08-30
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 9
;  LENGTH: 8933
;  TYPE: RNA
;  ORGANISM: Artificial Sequence
;  FEATURE:
;  OTHER INFORMATION: Sequence of the IAC-Bscr pseudo target
;  NAME/KEY: mutation
;  LOCATION: (4140)...(4159)
;  OTHER INFORMATION: Mutated positions: 4140-42, 4145-47, 4152,
;  OTHER INFORMATION: 4156-57, 4159
US-09-943-286-9

  Query Match             41.2%;  Score 1056.2;  DB 10;  Length 8933;
  Best Local Similarity   50.0%;  Pred. No. 2.8e-244;
  Matches 1258;  Conservative  353;  Mismatches  903;  Indels    3;  Gaps    1;

Qy       1 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 60
Db    1906 UUUCCCAUUAGCCCUAUUGAGACUGUACCAGUAAAAUUAAAGCCAGGAAUGGAUGGCCCA 1965

Qy      61 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 120
Db    1966 AAAGUUAAACAAUGGCCAUUGACAGAAGAAAAAAUAAAAGCAUUAGUAGAAAUUUGUACA 2025

Qy     121 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 180
Db    2026 GAAAUGGAAAAGGAAGGGAAAAUUUCAAAAAUUGGGCCUGAGAAUCCAUACAAUACUCCA 2085

Qy     181 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 240
Db    2086 GUAUUUGCCAUAAAGAAAAAAGACAGUACUAAAUGGAGAAAAUUAGUAGAUUUCAGAGAA 2145

Qy     241 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 300
Db    2146 CUUAAUAAGAGAACUCAAGACUUCUGGGAAGUUCAAUUAGGAAUACCACAUCCCGCAGGG 2205

Qy     301 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 360
Db    2206 UUAAAAAAGAAAAAAUCAGUAACAGUACUGGAUGUGGGUGAUGCAUAUUUUUCAGUUCCC 2265

Qy     361 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 420
Db    2266 UUAGAUGAAGACUUCAGGAAGUAUACUGCAUUUACCAUACCUAGUAUAAACAAUGAGACA 2325

Qy     421 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 480
Db    2326 CCAGGGAUUAGAUAUCAGUACAAUGUGCUUCCACAGGGAUGGAAAGGAUCACCAGCAAUA 2385

Qy     481 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 540
Db    2386 UUCCAAAGUAGCAUGACAAAAAUCUUAGAGCCUUUUAAAAAACAAAAUCCAGACAUAGUU 2445

Qy     541 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 600
Db    2446 AUCUAUCAAUACAUGGAUGAUUUGUAUGUAGGAUCUGACUUAGAAAUAGGGCAGCAUAGA 2505

Qy     601 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 660
Db    2506 ACAAAAAUAGAGGAGCUGAGACAACAUCUGUUCAGGUGGGGACUUACCACACCAGACAAA 2565

Qy     661 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 720
Db    2566 AAACAUCAGAAAGAACCUCCAUUCCUUUGGAUGGGUUAUGAACUCCAUCCUGAUAAAUGG 2625

Qy     721 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 780
```

```
Db   2626 ACAGUACAGCCUAUAGUGCUGCCAGAAAAAGACAGCUGGACUGUCAAUGACAUACAGAAG 2685

Qy    781 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 840

Db   2686 UUAGUGGGGAAAUUGAAUUGGGCAAGUCAGAUUUACCCAGGGAUUAAAGUAAGGCAAUUA 2745

Qy    841 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 900

Db   2746 UGUAAACUCCUUAGAGGAACCAAAGCACUAACAGAAGUAAUACCACUAACAGAAGAAGCA 2805

Qy    901 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 960

Db   2806 GAGCUAGAACUGGCAGAAAACAGAGAGAUUCUAAAAGAACCAGUACAUGGAGUGUAUUAU 2865

Qy    961 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 1020

Db   2866 GACCCAUCAAAAGACUUAAUAGCAGAAAUACAGAAGCAGGGGCAAGGCCAAUGGACAUAU 2925

Qy   1021 AAAATTCACCAAGA---AGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAAT 1077

Db   2926 CAAAUUUAUCAAGAGCCAUUUAAAAAUCUGAAAACAGGAAAAUAUGCAAGAAUGAGGGGU 2985

Qy   1078 ACACATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCA 1137

Db   2986 GCCCACACUAAUGAUGUAAAACAAUUAACAGAGGCAGUGCAAAAAAUAACCACAGAAAGC 3045

Qy   1138 ATAGTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAA 1197

Db   3046 AUAGUAAUAUGGGGAAAGACUCCUAAAUUUAAACUACCCAUACAAAAGGAAACAUGGGAA 3105

Qy   1198 CAGTGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACA 1257

Db   3106 ACAUGGUGGACAGAGUAUUGGCAAGCCACCUGGAUUCCUGAGUGGGAGUUUGUUAAUACC 3165

Qy   1258 CCACCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACC 1317

Db   3166 CCUCCUUUAGUGAAAUUAUGGUACCAGUUAGAGAAAGAACCCAUAGUAGGAGCAGAAACC 3225

Qy   1318 TATTATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACA 1377

Db   3226 UUCUAUGUAGAUGGGGCAGCUAACAGGGAGACUAAAUUAGGAAAAGCAGGAUAUGUUACU 3285

Qy   1378 GATAGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTG 1437

Db   3286 AACAAAGGAAGACAAAAGGUUGUCCCCCUAACUAACACAACAAAUCAGAAAACUGAGUUA 3345

Qy   1438 GAAGCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCA 1497

Db   3346 CAAGCAAUUUAUCUAGCUUUGCAGGAUUCAGGAUUAGAAGUAAACAUAGUAACAGACUCA 3405

Qy   1498 CAATATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAAT 1557

Db   3406 CAAUAUGCAUUAGGAAUCAUUCAAGCACAACCAGAUAAAAGUGAAUCAGAGUUAGUCAAU 3465

Qy   1558 CAAATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACAC 1617

Db   3466 CAAAUAAUAGAGCAGUUAAUAAAAAAGGAAAAGGUCUAUCUGGCAUGGGUACCAGCACAC 3525

Qy   1618 AAAGGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTT 1677

Db   3526 AAAGGAAUUGGAGGAAAUGAACAAGUAGAUAAAUUAGUCAGUGCUGGAAUCAGGAAAAUA 3585

Qy   1678 CTCTTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTA 1737

Db   3586 CUAUUUUUAGAUGGAAUAGAUAAGGCCCAAGAUGAACAUGAGAAAUAUCACAGUAAUUGG 3645

Qy   1738 AAAGAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACC 1797

Db   3646 AGAGCAAUGGCUAGUGAUUUUAACCUGCCACCUGUAGUAGCAAAAGAAAUAGUAGCCAGC 3705

Qy   1798 TGTGATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGG 1857

Db   3706 UGUGAUAAAUGUCAGCUAAAAGGAGAAGCCAUGCAUGGACAAGUAGACUGUAGUCCAGGA 3765

Qy   1858 ACTTGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTA 1917

Db   3766 AUAUGGCAACUAGAUUGUACACAUUUAGAAGGAAAAGUUAUCCUGGUAGCAGUUCAUGUA 3825

Qy   1918 GCTAGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTA 1977

Db   3826 GCCAGUGGAUAUAUAGAAGCAGAAGUUAUUCCAGCAGAAACAGGGCAGGAAACAGCAUAU 3885

Qy   1978 TTTCTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCT 2037

Db   3886 UUUCUUUUAAAAUUAGCAGGAAGAUGGCCAGUAAAAACAAUACAUACAGACAAUGGCAGC 3945

Qy   2038 AACTTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTT 2097

Db   3946 AAUUUCACCAGUGCUACGGUUAAGGCCGCCUGUUGGUGGGCGGGAAUCAAGCAGGAAUUU 4005

Qy   2098 GGGGTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAA 2157

Db   4006 GGAAUUCCCUACAAUCCCCAAAGUCAAGGAGUAGUAGAAUCUAUGAAUAAAGAAUUAAAG 4065

Qy   2158 AATCAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCA 2217

Db   4066 AAAAUUAUAGGACAGGUAAGAGAUCAGGCUGAACAUCUUAAGACAGCAGUACAAAUGGCA 4125

Qy   2218 GTTCATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGA 2277

Db   4126 GUAUUCAUCCAGAAAAAUAUUUGAAAGGGGAAGCUUGGGGGGUACAGUGCAGGGGAAAGA 4185

Qy   2278 TTAATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAA 2337

Db   4186 AUAGUAGACAUAAUAGCAACAGACAUACAAACUAAAGAAUUACAAAAACAAAUUACAAAA 4245

Qy   2338 TTTAAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGT 2397

Db   4246 AUUCAAAAUUUUCGGGUUUAUUACAGGGACAGCAGAAAUCCACUUUGGAAAGGACCAGCA 4305

Qy   2398 GAGCTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTA 2457

Db   4306 AAGCUCCUCUGGAAAGGUGAAGGGGCAGUAGUAAUACAAGAUAAUAGUGACAUAAAAGUA 4365

Qy   2458 GTACCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGAT 2514

Db   4366 GUGCCAAGAAGAAAAGCAAAGAUCAUUAGGGAUUAUGGAAAACAGAUGGCAGGUGAU 4422


RESULT 7
US-09-872-733-2
; Sequence 2, Application US/09872733
; Patent No. US20010036655A1
; GENERAL INFORMATION:
;  APPLICANT: The Government of the United States of America, as
;  TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, SIV GAG AND
;  TITLE OF INVENTION:  SIV ENV GENES
;  FILE REFERENCE: 2026-4287US1 HIV GAG/POL,SIV GAG & ENV
;  CURRENT APPLICATION NUMBER: US/09/872,733
;  CURRENT FILING DATE:  2001-06-01
;  PRIOR APPLICATION NUMBER: PCT/US00/34985
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/173,036
;  PRIOR FILING DATE: 1999-12-23
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 2507
;   TYPE: DNA
;   ORGANISM: Human immunodeficiency virus type 1
US-09-872-733-2

  Query Match             37.5%;  Score 961;  DB 10;  Length 2507;
  Best Local Similarity   63.0%;  Pred. No. 1.3e-221;
  Matches 1540;  Conservative    0;  Mismatches  860;  Indels   43;  Gaps    2;
```



```
Db   2161 TTTTAAAAGAAAAGGGGGGATTGGGGGGTACAGTGCAGGGGAAAGAATAGTAGACATAAT 2220

Qy   2292 CACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAATTTAAAAATTTTCG 2351
          | |||||   | | |   ||     ||||  | ||| ||||| || ||||||||||
Db   2221 AGCAACAGACATACAAACTAAAGAATTACAAAAACAAATTACAAAAATTCAAAATTTTCG 2280

Qy   2352 GGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGTGAGCTATTGTGGAA 2411
          ||| |||||||| ||  ||||| ||| ||| ||||| ||||| |   ||||  | |||||
Db   2281 GGTTTATTACAGGGACAGCAGAAATCCACTTTGGAAAGGACCAGCAAAGCTCCTCTGGAA 2340

Qy   2412 AGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTAGTACCCAGAAGAAA 2471
          ||| ||||| ||||| |  || | |     |  ||||| || ||||| || ||||||||
Db   2341 AGGTGAAGGGGCAGTAGTAATACAAGATAATAGTGACATAAAAGTAGTGCCAAGAAGAAA 2400

Qy   2472 GGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGAT 2514
           || || || || |  |||||||||  |   |  |  |  |||
Db   2401 AGCAAAGATCATTAGGGATTATGGAAAACAGATGGCAGGTGAT 2443


RESULT 8
US-10-097-997-1
; Sequence 1, Application US/10097997
; Patent No. US20020160406A1
; GENERAL INFORMATION:
;  APPLICANT: Abbott Laboratories
;  APPLICANT:  Salituro, John A
;  APPLICANT:  Carrino, John J.
;  TITLE OF INVENTION: Method For Using Unequal Primer
;  TITLE OF INVENTION:  Concentrations For Generating Nucleic Acid Amplification
;  TITLE OF INVENTION:  Products
;  FILE REFERENCE: 6337.US.C1
;  CURRENT APPLICATION NUMBER: US/10/097,997
;  CURRENT FILING DATE:  2002-03-14
;  PRIOR APPLICATION NUMBER: US 09/079,675
;  PRIOR FILING DATE: 1998-05-15
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 2348
;   TYPE: DNA
;   ORGANISM: HIV-1
US-10-097-997-1

  Query Match             37.4%;  Score 958.4;  DB 9;  Length 2348;
  Best Local Similarity   64.2%;  Pred. No. 5.2e-221;
  Matches 1458;  Conservative    0;  Mismatches  811;  Indels    3;  Gaps    1;

Qy      1 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 60
          ||||||||   |    | ||  |||||  |||    ||||||||||||| |||||| |||
Db     76 TTTCCCATTAGTCCTATTGAAACTGTACCAGTAAAATTAAAGCCAGGAATGGATGGCCCA 135

Qy     61 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 120
          ||| | || || |||||||| || |||||||.|||   |||||| |||||| ||| |
Db    136 AAAGTTAAACAATGGCCATTGACAGAAGAAAAAATAAAAGCATTAATAGAAATTTGTACA 195

Qy    121 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 180
          | |||||||||||| ||  |  |      ||. |   ||    |||||||||| || ||
Db    196 GAAATGGAAAAGGAAGGGAAAATTTCAAAAATTGGGCCTGAAAATCCATACAATACTCCA 255

Qy    181 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 240
          ||||||  |||||||||||| || |  |  |||||||||    |  ||||||| || |||
Db    256 GTATTTGCCATAAAGAAAAAAGACAGTACTAAATGGAGAAAATTAGTAGATTTCAGAGAA 315

Qy    241 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 300
          || |||| |   |||||  |||||   |||||| |||||||||||||||| ||||||||
Db    316 CTTAATAAGAAAACTCAAGACTTCTGGGAAGTTCAATTAGGAATACCACATCCTGCAGGG 375

Qy    301 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 360
          ||  ||| |  |||  | | ||||||||||||| | ||||||||||||||| ||  | ||
Db    376 TTAAAAAAGAAAAAATCAGTAACAGTACTGGATGTGGGTGATGCATATTTTTCAGTTCCC 435

Qy    361 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 420
          ||||| |||| || ||| |||| ||||| |||||  ||||     |||| |||| |

Db    436 TTAGATAAAGACTTCAGGAAGTATACTGCATTTACCATACCTAGTATAAACAATGAAACA 495

Qy    421 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 480
          ||||| |   ||||    || || || || || ||||||||||| || ||||||||| ||
Db    496 CCAGGGATTAGATATCAGTACAATGTGCTTCCACAGGGATGGAAAGGATCACCAGCAATA 555

Qy    481 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 540
          ||||||   |  |||| | |  |  |||| || || || ||   |||||||||  ||
Db    556 TTCCAAAGTAGCATGACAAAAATCTTAGAGCCTTTTAGAAAACAAAATCCAGACATAGTT 615

Qy    541 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 600
          |   || || |||||||| | |  |||   ||||    |    | | |||
Db    616 ATCTATCAATACATGGATGATTTGTATGTAGGATCTGACTTAGAAATAGGGCAGCATAGA 675

Qy    601 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 660
          | |     |   |    ||  ||  ||    || ||| | ||  |||||  |
Db    676 GCAAAAATAGAGGAACTGAGACGACATCTGTTGAGGTGGGGATTTACCACACCAGACAAA 735

Qy    661 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 720
          |||   || ||||| || ||||| |  |||||||| || ||| |    ||    ||||||
Db    736 AAACATCAGAAAGAACCTCCATTCCTTTGGATGGGTTATGAACTCCATCCTGATAAATGG 795

Qy    721 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 780
          |   | ||    |||| | | ||| ||| ||| | |||||| || ||||| |||||||||
Db    796 ACAGTACAGCCTATAGTGCTACCAGAAAAAGACAGCTGGACTGTCAATGACATACAGAAG 855

Qy    781 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 840
          ||||| |||  ||| |||||||||  ||| |||||. |||| || |||   || ||  |
Db    856 TTAGTGGGAAAATTGAATTGGGCAAGTCAGATTTACGCAGGGATTAAAGTAAAGCAATTA 915

Qy    841 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 900
          ||||   |||||||||  |   | |||||||.| | | |    || ||   |||
Db    916 TGTAAACTCCTTAGAGGAACCAAAGCACTAACAGAAGTAATACCACTAACAGAAGAAGCA 975

Qy    901 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 960
          ||   ||||   | ||||| |   |||||| |  |   | | | |||   |||||
Db    976 GAGCTAGAACTGGCAGAAAACAGGGAAATTCTAAAAGAACCAGTACATGGAGTGTATTAT 1035

Qy    961 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 1020
          |   |  ||   ||| |||   |||  |||     | |    || ||| | |||
Db   1036 GACCCATCAAAAGACTTAATAGCAGAAGTACAGAAGCAGGGGCAAGGCCAATGGACATAT 1095

Qy   1021 AAAATTCACCAAGA---AGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAAT 1077
          ||||| | ||||    |  ||||  ||||||  ||| |||| |||||  || | |  |
Db   1096 CAAATTTATCAAGAGCCATTTAAAAATCTGAAAACAGGCAAATATGCAAGAATGAGGGGT 1155

Qy   1078 ACACATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCA 1137
          | || || ||||  || | || |||| || | | | | || |||||||  |  |||
Db   1156 GCCCACACTAATGATGTAAAACAATTAACAGAGGCAGTGCAAAAAATAGCCACAGAAAGC 1215

Qy   1138 ATAGTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAA 1197
          ||||| || |||||| ||   || |||||     |||| |  | || ||   |||||||
Db   1216 ATAGTAATATGGGGAAAGACTCCTAAATTTAGACTACCCATACAAAAAGAAACATGGGAA 1275

Qy   1198 CAGTGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACA 1257
             |||||||||| ||| | | |||||||| || || ||||  || ||  ||
Db   1276 ACATGGTGGACAGAGTATACGTAAGCCACCTGGATTCCTGAGTGGGAGGTTGTCAATACC 1335

Qy   1258 CCACCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACC 1317
          || ||  |||| | |||   | |  ||| ||| || || |||   || | ||||||
Db   1336 CCTCCCTTAGTGAAATTATGGTACCAGTTAGAGAAAGAACCCATAGTAGGTGCAGAAACT 1395

Qy   1318 TATTATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACA 1377
          |  |||  ||||||  ||  ||| |   || | ||| |||| ||||||||||||| || ||
Db   1396 TTCTATGTAGATGGGGCAGCTAACAGGGAGACTAAAAAAGGAAAAGCAGGATATGTTACT 1455

Qy   1378 GATAGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTG 1437
          | || || | | | ||  ||    |-|| || ||||  |  | || ||
Db   1456 AACAGAGGAAGACAAAAGGTTGTCTCCCTAACTGACACAACAAATCAGAAGACTGAGTTA 1515

Qy   1438 GAAGCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCA 1497
          ||||| |||   | || |||  ||| ||||||  | | | ||| ||  ||  ||| |||
Db   1516 CAAGCAATTCATCTAGCTTTGCAAGATTCAGGGTTAGAAGTAAACATAGTAACAGACTCA 1575
```

```
Qy   1498 CAATATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAAT 1557
          |||||||   | |||||  ||   ||  |   ||     ||   ||       | ||||  | |
Db   1576 CAATATGCATTAGGAATCATTCAAGCACAACCAGATAAAAGTGAATCAGAGTTAGTCAGT 1635

Qy   1558 CAAATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACAC 1617
          ||||||||||   |  | ||  ||||||  | |  | ||| | ||||||||||||||||||
Db   1636 CAAATAATAGAGCAGTTAATAAAAAAGGAAAAGGTCTATCTGGCATGGGTACCAGCACAC 1695

Qy   1618 AAAGGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTT 1677
          |||||  ||  ||||||||  ||  ||  ||||  |  ||||  |||   ||  ||  ||  ||||
Db   1696 AAAGGAATTGGAGGAAATGAACAAGTAGATAAATTAGTCAGTGCTGGAATCAGGAAAGTA 1755

Qy   1678 CTCTTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTA 1737
          ||  ||  ||  ||    |||||    ||  ||||||||  |||||  |||||  ||  ||||||
Db   1756 CTATTTTTAGATGGAATAGATAAGGCCCAAGAAGACCATGAGAAATATCACAGTAATTGG 1815

Qy   1738 AAAGAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACC 1797
          |  ||  |  |||      | |||    ||||  |||||  ||  |  |  |  |||||||  ||  |
Db   1816 AGAGCAATGGCTAGTGACTTTAACCTACCACCTATAGTAGCAAAAGAAATAGTAGCCAGC 1875

Qy   1798 TGTGATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGG 1857
          ||||||||||||||  |  ||||||||  ||  ||  |||||  ||  |  |  |  |    ||  |||
Db   1876 TGTGATAAATGTCAGCTAAAAGGAGAAGCCATGCATGGACAAGTAGACTGTAGTCCAGGA 1935

Qy   1858 ACTTGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTA 1917
          |  ||||||  |  ||||||||  |||  ||||  ||||||  |  |||  |  ||  |||||  ||||||
Db   1936 ATATGGCAACTAGATTGTACACATTTAGAAGGAAAAGTTATCCTGGTAGCAGTTCATGTA 1995

Qy   1918 GCTAGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTA 1977
          ||  |||||||  ||||||||||||  ||  ||||||  |||||||||     |||||||||
Db   1996 GCCAGTGGATACATAGAAGCAGAAGTTATTCCAGCAGAGACAGGGCAGGAGACAGCATAC 2055

Qy   1978 TTTCTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCT 2037
          |||||  ||||||||  |||||  ||||||||  |  |  |    ||||  |||||  |||||  |
Db   2056 TTTCTCTTAAAATTAGCAGGAAGATGGCCAGTAAAAACAATACATACAGACAATGGCCCC 2115

Qy   2038 AACTTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTT 2097
          ||  ||  |          ||  |||    |      ||||||  |  ||  ||  ||||    |||
Db   2116 AATTTCACCAGTACTACGGTTAAGGCCGCCTGTTGGTGGACGGGAATCAAGCAGGAATTT 2175

Qy   2098 GGGGTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAA 2157
          ||  |  ||  ||||||||  ||  |||||  ||||||  |  |||  |  ||||||  |   |  |  ||
Db   2176 GGCATTCCCTACAATCCCCAAAGTCAAGGAGTAATAGAATCTATGAATAAAGAATTAAAG 2235

Qy   2158 AATCAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCA 2217
          ||    ||||      | ||  ||  ||  ||  |     |  |  |    ||||  |||||||
Db   2236 AAAATTATAGGACAGGTAAGAGATCAGGCTGAACATCTTAAGAGAGCAGTACAAATGGCA 2295

Qy   2218 GTTCATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAG 2269
          ||      |   ||||||||||||||  ||  ||  ||  ||||     |  |  |||
Db   2296 GTATTCATCCACAATTTTAAAAGAAAAGGGGGGATTGGGGGGTACAGTGCAG 2347


RESULT 9
US-09-872-733-6
; Sequence 6, Application US/09872733
; Patent No. US20010036655A1
; GENERAL INFORMATION:
;  APPLICANT: The Government of the United States of America, as
;  TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, SIV GAG AND
;  TITLE OF INVENTION:  SIV ENV GENES
;  FILE REFERENCE: 2026-4287US1 HIV GAG/POL,SIV GAG & ENV
;  CURRENT APPLICATION NUMBER: US/09/872,733
;  CURRENT FILING DATE:  2001-06-01
;  PRIOR APPLICATION NUMBER: PCT/US00/34985
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/173,036
;  PRIOR FILING DATE: 1999-12-23
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 8366
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  DNA sequence
;   OTHER INFORMATION: of the construct pCMVgagpolBNKan containing a CMV
;   OTHER INFORMATION: promoter, a HIV gag/pol gene and a kanamycin
;   OTHER INFORMATION: resistance gene
US-09-872-733-6

  Query Match             34.6%;  Score 888.2;  DB 10;  Length 8366;
  Best Local Similarity   59.8%;  Pred. No. 6.9e-204;
  Matches 1506;  Conservative    0;  Mismatches 1008;  Indels    3;  Gaps    1;

Qy      1 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 60
          ||  |||||         | |||  | ||    ||    ||  |||||  ||  |  |||  ||  ||
Db   2527 TTCCCCATCAGCCCTATTGAGACGGTGCCCGTGAAGTTGAAGCCGGGGATGGACGGCCCC 2586

Qy     61 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 120
          ||  |  |||||  |||||||  |  |||||  |||||    ||  |||   |||||||||  |
Db   2587 AAGGTCAAGCAATGGCCATTGACGAAAGAGAAGATCAAGGCCTTAGTCGAAATCTGTACA 2646

Qy    121 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 180
          |||||||  |||||  ||  ||  |      |        ||    ||  ||  ||||||||  ||
Db   2647 GAGATGGAGAAGGAAGGGAAGATCAGCAAGATCGGGCCTGAGAACCCCTACAACACTCCA 2706

Qy    181 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 240
            ||  ||  ||  |||||  |||||  |   |  |||  |||||||  ||||  |  ||  ||  ||
Db   2707 GTCTTCGCAATCAAGAAGAAGGACAGTACCAAGTGGAGAAAGCTGGTGGACTTCAGAGAG 2766

Qy    241 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 300
          ||  ||  |  |  ||||||||||   ||||||  ||   |  ||  ||  |||||  ||  ||
Db   2767 CTGAACAAGAGAACTCAGGACTTCTGGGAAGTTCAGCTGGGCATCCCACATCCCGCTGGG 2826

Qy    301 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 360
            |     ||  |  |||    |  |  |||||  ||||||  |  ||||||||  ||  ||||||  |  ||
Db   2827 TTGAAGAAGAAGAAGTCAGTGACAGTGCTGGATGTGGGTGATGCCTACTTCTCCGTTCCC 2886

Qy    361 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 420
            |  ||  ||  ||  ||  |||  ||||||||||||||  ||  ||||      |  ||  ||  |
Db   2887 TTGGACGAGGACTTCAGGAAGTACACTGCCTTCACGATACCTAGCATCAACAACGAGACA 2946

Qy    421 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 480
          |||||  |   ||  |||    ||  ||  |||||  ||||||||||||  ||||||||||||
Db   2947 CCAGGCATCCGCTACCAGTACAACGTGCTGCCACAGGGATGGAAGGGATCACCAGCCATC 3006

Qy    481 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 540
          ||  |||  ||  ||||    |  |  ||  ||  |||||  |  |||  |||  |||||  |
Db   3007 TTTCAAAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAAAACCCAGACATCGTG 3066

Qy    541 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 600
            |    |||||  |||||  |||  |||    |||  ||||||  |      |  |  |  ||
Db   3067 ATCTATCAGTACATGGACGACCTCTACGTAGGAAGTGACCTGGAGATCGGGCAGCACAGG 3126

Qy    601 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 660
          |       |     ||   |  |  |  ||   ||     ||  |  |  |  ||  |||||  ||
Db   3127 ACCAAGATCGAGGAGCTGAGACAGCATCTGTTGAGGTGGGGACTGACCACACCAGACAAG 3186

Qy    661 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 720
          ||    |||  ||  ||  ||  ||  ||  |  ||||||||  ||||||  ||    ||     ||  |||
Db   3187 AAGCACCAGAAGGAACCTCCCTTCCTGTGGATGGGCTACGAACTGCATCCTGACAAGTGG 3246

Qy    721 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 780
            |   ||||     ||  |  |  ||||  |  |   ||  |  ||||||  |||||  ||  ||||||||||
Db   3247 ACAGTGCAGCCCATCGTGCTGCCTGAGAAGGACAGCTGGACTGTGAACGACATACAGAAG 3306

Qy    781 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 840
            |  ||  ||     ||  ||  ||||||    ||  ||  ||  |||||  ||  |||     |   ||  ||
Db   3307 CTCGTGGGCAAGTTGAACTGGGCAAGCCAGATCTACCCAGGCATCAAAGTTAGGCAGCTG 3366

Qy    841 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 900
          ||  |  |  |  ||  ||||||  |  |  |  ||  |||||  |    |  |     |||  |||    |||
Db   3367 TGCAAGCTGCTTCGAGGAACCAAGGCACTGACAGAAGTGATCCCACTGACAGAGGAAGCA 3426
```

```
Qy    901 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 960
          ||   ||||   |  || ||   |   || || |   ||  |  | | |||   || |||
Db   3427 GAGCTAGAACTGGCAGAGAACCGAGAGATCCTGAAGGAGCCAGTACATGGAGTGTACTAC 3486

Qy    961 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 1020
          |   | |||||    |    ||   | |    |||      | |    || ||| | ||
Db   3487 GACCCAAGCAAGGACCTGATCGCAGAGATCCAGAAGCAGGGGCAAGGCCAATGGACCTAC 3546

Qy   1021 AAAATTCACCAAGA---AGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAAT 1077
          ||||  |||| ||      ||| |  |||||   ||| || |  |||| ||| | |  |
Db   3547 CAAATCTACCAGGAGCCCTTCAAGAACCTGAAGACAGGCAAGTACGCAAGGATGAGGGGT 3606

Qy   1078 ACACATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCA 1137
          | || ||||| |  |||| |   |  || | | | | |||| ||    |  ||
Db   3607 GCCCACACCAACGATGTGAAGCAGCTGACAGAGGCAGTGCAGAAGATCACCACAGAGAGC 3666

Qy   1138 ATAGTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAA 1197
          || ||||||||||||  ||   || || ||| |  | ||  |  |||||||   ||||||
Db   3667 ATCGTGATCTGGGGCAAGACTCCCAAGTTCAAGCTGCCCATACAGAAGGAGACATGGGAG 3726

Qy   1198 CAGTGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACA 1257
          ||||||||| || || ||||| |  |||||||| || || ||||| ||  |    ||
Db   3727 ACATGGTGGACCGAGTACTGGCAAGCCACCTGGATCCCTGAGTGGGAGTTCGTGAACACC 3786

Qy   1258 CCACCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACC 1317
          || ||  | || | |    |  || | |||||| || || | ||||| ||| |||
Db   3787 CCTCCCTTGGTGAAACTGTGGTATCAGCTGGAGAAGGAACCCATCGTGGGAGCAGAGACC 3846

Qy   1318 TATTATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACA 1377
          |  ||    |||||  ||   || |   || |.||  ||  || |||||  ||  | ||
Db   3847 TTCTACGTGGATGGGGCAGCCAACAGGGAGACCAAGCTGGGCAAGGCAGGCTACGTGACC 3906

Qy   1378 GATAGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTG 1437
          |  | ||   | | |||||         |  . | || || || ||  |  | ||  ||
Db   3907 AACCGAGGACGACAGAAAGTGGTGACCCTGACTGACACCACCAACCAGAAGACTGAGCTG 3966

Qy   1438 GAAGCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCA 1497
          ||||  |   | | ||  ||  ||||   || |   | |  || ||  |   ||| |||
Db   3967 CAAGCCATCTACCTAGCTCTGCAAGACAGCGGACTGGAAGTGAACATCGTGACAGACTCA 4026

Qy   1498 CAATATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAAT 1557
          || || |   |||| || ||   || |   ||      |||| |||     ||| || ||
Db   4027 CAGTACGCACTGGGCATCATCCAAGCACAACCAGACCAATCCGAGTCAGAGCTGGTGAAC 4086

Qy   1558 CAAATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACAC 1617
          || || || ||  |  |||| || |||    || | ||  |.||||||||||||||||||
Db   4087 CAGATCATCGAGCAGCTGATCAAGAAGGAGAAAGTGTACCTGGCATGGGTACCAGCACAC 4146

Qy   1618 AAAGGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTT 1677
          ||||| || |||||||  || || |||  |  |||| |||  |||||  |  | ||
Db   4147 AAAGGAATTGGAGGAAATGAACAAGTAGATAAATTAGTCAGTGCTGGGATCCGGAAGGTG 4206

Qy   1678 CTCTTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTA 1737
          || ||| ||||   ||| ||    || ||||| |||||||| || |||||    ||
Db   4207 CTGTTCCTGGACGGGATCGATAAGGCCCAAGATGAACATGAGAAGTACCACTCCAACTGG 4266

Qy   1738 AAAGAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACC 1797
          |   |||    | | ||     | ||     |||| || | | | ||||||| || |
Db   4267 CGCGCTATGGCCAGCGACTTCAACCTGCCACCTGTAGTAGCAAAAGAAATAGTAGCCAGC 4326

Qy   1798 TGTGATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGG 1857
          |||||||||||||  |  |||||||| || || ||||| || | | | |    || |||
Db   4327 TGTGATAAATGTCAGCTAAAAGGAGAAGCCATGCATGGACAAGTAGACTGTAGTCCAGGA 4386

Qy   1858 ACTTGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTA 1917
          |  |||||  |||| || || || || ||||| ||  | ||| | || ||||| ||||||
Db   4387 ATATGGCAGCTGGACTGCACGCACCTGGAGGGGAAGGTGATCCTGGTAGCAGTTCATGTA 4446

Qy   1918 GCTAGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTA 1977
          || |||||||  |||||||||||| || || ||    || || ||      | ||||||
Db   4447 GCCAGTGGATATATAGAAGCAGAAGTTATCCCTGCTGAAACTGGGCAGGAAACAGCATAT 4506

Qy   1978 TTTCTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCT 2037
          |||||  |||||||| |||||  |||||||| | | |    ||||||| || || ||
Db   4507 TTTCTTTTAAAATTAGCAGGAAGATGGCCAGTAAAAACAATACACACGGACAACGGAAGC 4566

Qy   2038 AACTTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTT 2097
          |||||  ||          || |||   |     |||||||| || ||  ||||   |||
Db   4567 AACTTCACTGGTGCTACGGTTAAGGCCGCCTGTTGGTGGGCGGGAATCAAGCAGGAATTT 4626

Qy   2098 GGGGTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAA 2157
          ||  | || ||||||||| ||     || |||||  |||||     |||||  |  | |||||
Db   4627 GGAATTCCCTACAATCCCCAATCGCAAGGAGTCGTGGAGAGCATGAACAAGGAGCTGAAG 4686

Qy   2158 AATCAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCA 2217
          ||     || |    | | ||||| || ||  |   |  | ||   |||    ||||||
Db   4687 AAGATCATCGGACAAGTGAGGGATCAGGCTGAGCACCTGAAGACAGCAGTGCAGATGGCA 4746

Qy   2218 GTTCATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGA 2277
          ||       |   || || |||||||  || || || ||||     | | |||  |||||
Db   4747 GTGTTCATCCACAACTTCAAAAGAAAAGGGGGGATTGGGGGGTACAGTGCAGGGGAAAGG 4806

Qy   2278 TTAATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAA 2337
          |  |  |||| ||| | || ||      | |   |    |  ||      | | | | ||
Db   4807 ATCGTGGACATCATCGCCACCGACATCCAAACCAAGGAGCTGCAGAAGCAGATCACCAAG 4866

Qy   2338 TTTAAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGT 2397
          |  | || || ||||| || ||| | ||  || |  | | ||||||||||||||| |
Db   4867 ATCCAGAACTTCCGGGTGTACTACCGCGACAGCCGCAACCCACTGTGGAAGGGACCAGCA 4926

Qy   2398 GAGCTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTA 2457
          ||||  | ||||| || || || |||||  |  |  |||      |  ||||| || ||
Db   4927 AAGCTCCTCTGGAAGGGAGAGGGGGCAGTGGTGATCCAGGACAACAGTGACATCAAAGTG 4986

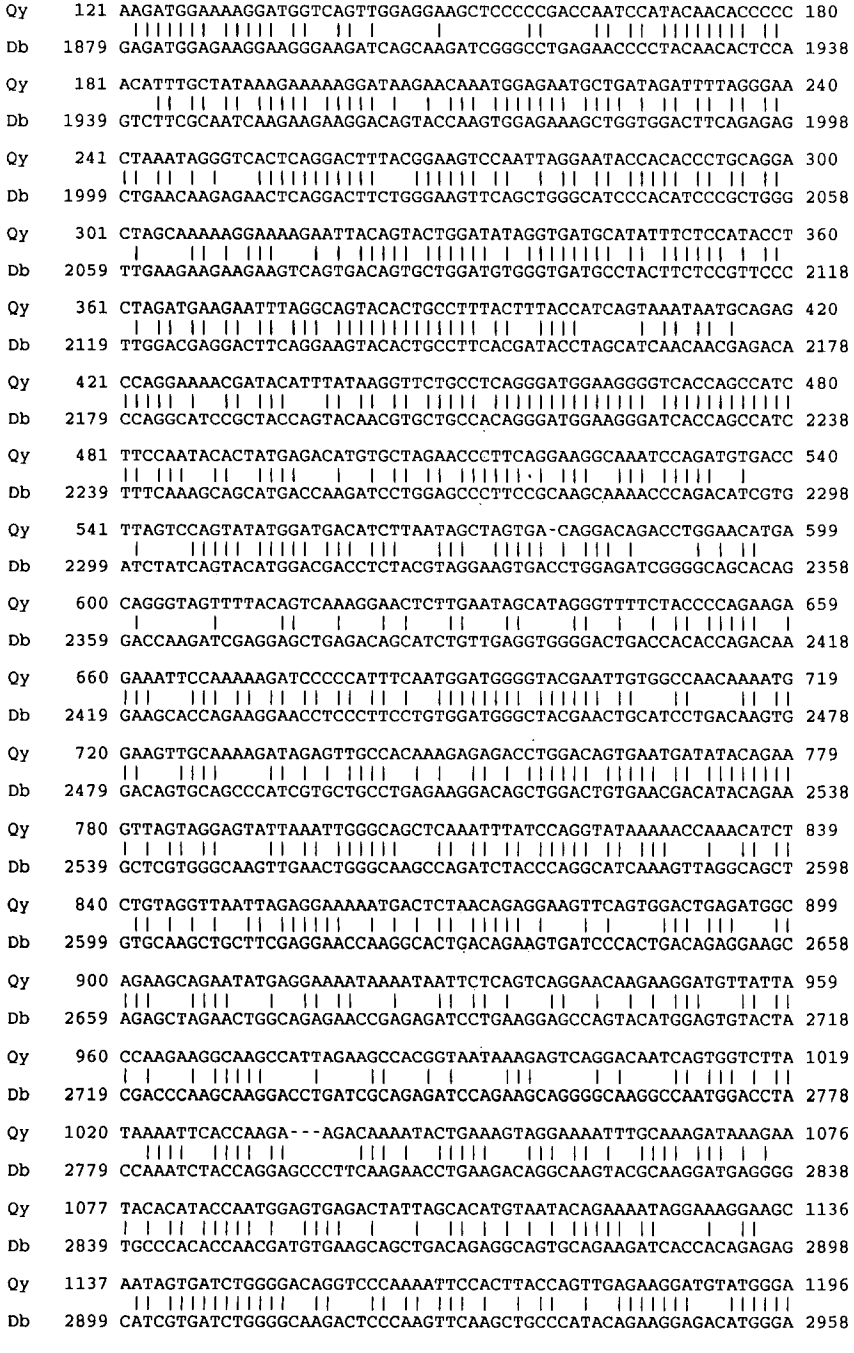
Qy   2458 GTACCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGAT 2514
          || || ||  | ||||| || || |||   || ||||||  |   |  |  |  |||
Db   4987 GTGCCAAGGCGCAAGGCCAAGATCATCCGCGACTATGGAAAACAGATGGCAGGTGAT 5043


RESULT 10
US-09-872-733-1
; Sequence 1, Application US/09872733
; Patent No. US20010036655A1
; GENERAL INFORMATION:
;  APPLICANT: The Government of the United States of America, as
;  TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, SIV GAG AND
;  TITLE OF INVENTION:  SIV ENV GENES
;  FILE REFERENCE: 2026-4287US1 HIV GAG/POL,SIV GAG & ENV
;  CURRENT APPLICATION NUMBER: US/09/872,733
;  CURRENT FILING DATE:  2001-06-01
;  PRIOR APPLICATION NUMBER: PCT/US00/34985
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/173,036
;  PRIOR FILING DATE: 1999-12-23
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;  LENGTH: 4338
;  TYPE: DNA
;  ORGANISM: Artificial Sequence
;  FEATURE:
;  OTHER INFORMATION: Description of Artificial Sequence:  Mutated Human
;  OTHER INFORMATION: Immunodeficiency Virus - 1 Gag/Pol gene
US-09-872-733-1

  Query Match             33.7%;  Score 862.8;  DB 10;  Length 4338;
  Best Local Similarity   60.0%;  Pred. No. 6.6e-198;
  Matches 1510;  Conservative    0;  Mismatches 1002;  Indels    6;  Gaps    4;

Qy       1 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 60
           || |||||         | ||| | ||     ||  ||||| || | ||| || ||
Db    1759 TTCCCCATCAGCCCTATTGAGACGGTGCCCGTGAAGTTGAAGCCGGGGATGGACGGCCCC 1818

Qy      61 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 120
           ||  | |||||  ||||||||  | |||||  |||||     || |||    |||||||||  |
Db    1819 AAGGTCAAGCAATGGCCATTGACGAAAGAGAAGATCAAGGCCTTAGTCGAAATCTGTACA 1878
```


Qy 121 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 180
Db 1879 GAGATGGAGAAGGAAGGGAAGATCAGCAAGATCGGGCCTGAGAACCCCTACAACACTCCA 1938
Qy 181 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 240
Db 1939 GTCTTCGCAATCAAGAAGAAGGACAGTACCAAGTGGAGAAAGCTGGTGGACTTCAGAGAG 1998
Qy 241 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 300
Db 1999 CTGAACAAGAGAACTCAGGACTTCTGGGAAGTTCAGCTGGGCATCCCACATCCCGCTGGG 2058
Qy 301 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 360
Db 2059 TTGAAGAAGAAGAAGTCAGTGACAGTGCTGGATGTGGGTGATGCCTACTTCTCCGTTCCC 2118
Qy 361 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 420
Db 2119 TTGGACGAGGACTTCAGGAAGTACACTGCCTTCACGATACCTAGCATCAACAACGAGACA 2178
Qy 421 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 480
Db 2179 CCAGGCATCCGCTACCAGTACAACGTGCTGCCACAGGGATGGAAGGGATCACCAGCCATC 2238
Qy 481 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 540
Db 2239 TTTCAAAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAAAACCCAGACATCGTG 2298
Qy 541 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGA-CAGGACAGACCTGGAACATGA 599
Db 2299 ATCTATCAGTACATGGACGACCTCTACGTAGGAAGTGACCTGGAGATCGGGGCAGCACAG 2358
Qy 600 CAGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGA 659
Db 2359 GACCAAGATCGAGGAGCTGAGACAGCATCTGTTGAGGTGGGGACTGACCACACCAGACAA 2418
Qy 660 GAAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATG 719
Db 2419 GAAGCACCAGAAGGAACCTCCCTTCCTGTGGATGGGCTACGAACTGCATCCTGACAAGTG 2478
Qy 720 GAAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAA 779
Db 2479 GACAGTGCAGCCCATCGTGCTGCCTGAGAAGGACAGCTGGACTGTGAACGACATACAGAA 2538
Qy 780 GTTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCT 839
Db 2539 GCTCGTGGGCAAGTTGAACTGGGCAAGCCAGATCTACCCAGGCATCAAAGTTAGGCAGCT 2598
Qy 840 CTGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGC 899
Db 2599 GTGCAAGCTGCTTCGAGGAACCAAGGCACTGACAGAAGTGATCCCACTGACAGAGGAAGC 2658
Qy 900 AGAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTA 959
Db 2659 AGAGCTAGAACTGGCAGAGAACCGAGAGATCCTGAAGGAGCCAGTACATGGAGTGTACTA 2718
Qy 960 CCAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTA 1019
Db 2719 CGACCCAAGCAAGGACCTGATCGCAGAGATCCAGAAGCAGGGGCAAGGCCAATGGACCTA 2778
Qy 1020 TAAAATTCACCAAGA---AGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAA 1076
Db 2779 CCAAATCTACCAGGAGCCCTTCAAGAACCTGAAGACAGGCAAGTACGCAAGGATGAGGGG 2838
Qy 1077 TACACATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGC 1136
Db 2839 TGCCCACACCAACGATGTGAAGCAGCTGACAGAGGCAGTGCAGAAGATCACCACAGAGAG 2898
Qy 1137 AATAGTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGA 1196
Db 2899 CATCGTGATCTGGGGCAAGACTCCCAAGTTCAAGCTGCCCATACAGAAGGAGACATGGGA 2958
Qy 1197 ACAGTGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAAC 1256
Db 2959 GACATGGTGGACCGAGTACTGGCAAGCCACCTGGATCCCTGAGTGGGAGTTCGTGAACAC 3018
Qy 1257 ACCACCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAAC 1316
Db 3019 CCCTCCCTTGGTGAAACTGTGGTATCAGCTGGAGAAGGAACCCATCGTGGGAGCAGAGAC 3078
Qy 1317 CTATTATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCAC 1376
Db 3079 CTTCTACGTGGATGGGGCAGCCAACAGGGAGACCAAGCTGGGCAAGGCAGGCTACGTGAC 3138
Qy 1377 AGATAGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATT 1436
Db 3139 CAACCGAGGACGACAGAAAGTGGTGACCCTGACTGACACCACCAACCAGAAGACTGAGCT 3198
Qy 1437 GGAAGCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTC 1496
Db 3199 GCAAGCCATCTACCTAGCTCTGCAAGACAGCGGACTGGAAGTGAACATCGTGACAGACTC 3258
Qy 1497 ACAATATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAA 1556
Db 3259 ACAGTACG-CATGGGCATCATCCAAGCACAACCAGACCAATCCGAGTCAGAGCTGGTGAA 3317
Qy 1557 TCAAATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACA 1616
Db 3318 CCAGATCATCGAGCAGCTGATCAAGAAGGAGAAAGTGTACCTGGCATGGGTACCAGCACA 3377
Qy 1617 CAAAGGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGT 1676
Db 3378 CAAAGGAATTGGAGGAAATGAACAAGTAGATAAATTAGTCAGTGCTGGGATCCGGAAGGT 3437
Qy 1677 TCTCTTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGT 1736
Db 3438 GCTGTTCCTGGACGGGATCGATAAGGCCCAAGATGAACATGAGAAGTACCACTCCAACTG 3497
Qy 1737 AAAAGAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACAC 1796
Db 3498 GCGCGCTATGGCCAGCGACTTCAACCTGCCACCTGTAGTAGCAAAAGAAATAGTAGCCAG 3557
Qy 1797 CTGTGATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGG 1856
Db 3558 CTGTGATAAATGTCAGCTAAAAGGAGAAGCCATGCATGGACAAGTAGACTGTAGTCCAGG 3617
Qy 1857 GACTTGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGT 1916
Db 3618 AATATGGCAGCTGGACTGCACGCACCTGGAGGGGAAGGTGATCCTGGTAGCAGTTCATGT 3677
Qy 1917 AGCTAGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACT 1976
Db 3678 AGCCAGTGGATATATAGAAGCAGAAGTTATCCCTGCTGAAACTGGGCAGGAAACAGCATA 3737
Qy 1977 ATTTCTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGC 2036
Db 3738 TTTTCTTTTAAAATTAGCAGGAAGATGGCCAGTAAAAACAATACACACGGACAACGGAAG 3797
Qy 2037 TAACTTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTT 2096
Db 3798 CAACTTCACTGGTGCTACGGTTAAGGCCGCCTGTTGGTGGGCGGGAATCAAGCAGGAATT 3857
Qy 2097 TGGGGTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAA 2156
Db 3858 TGGAATTCCCTACAATCCCCAATCGCAAGGAGTCGTGGAGAGCATGAACAAGGAGCTGAA 3917
Qy 2157 AAATCAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGC 2216
Db 3918 GAAGATCATCG-GACAGTGAGGGATCAGGCTGAGCACCTGAAGACAGCAGTGCAGATGGC 3976
Qy 2217 AGTTCATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAG 2276
Db 3977 AGTGTTCATCCACAACTTCAAAAGAAAAGGGGGGATTGGGGGGTACAGTGCAGGGGAAAG 4036
Qy 2277 ATTAATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAA 2336

```
Db   4037 GATCGTGGACATCATCGCCACCGACATCCAAACCAAGGAGCTGCAGAAGCAGATCACCAA 4096

Qy   2337 ATTTAAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGG 2396

Db   4097 GATCCAGAACTTCCGGGTGTACTACCGCGACAGCCGCAACCCACTGTGGAAGGGACCAGC 4156

Qy   2397 TGAGCTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGT 2456

Db   4157 AAAGCTCCTCTGGAAGGGAGAGGGGGCAGTGGTGATCCAGGACAACAGTGACATCAAAGT 4216

Qy   2457 AGTACCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGAT 2514

Db   4217 GGTGCCAAGGCGCAAGGCCAAGATCATCCGCGACTATGGAAAACAGATGGCAGGTGAT 4274


RESULT 11
US-09-872-733-3
; Sequence 3, Application US/09872733
; Patent No. US20010036655A1
; GENERAL INFORMATION:
;  APPLICANT: The Government of the United States of America, as
;  TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, SIV GAG AND
;  TITLE OF INVENTION:  SIV ENV GENES
;  FILE REFERENCE: 2026-4287US1 HIV GAG/POL,SIV GAG & ENV
;  CURRENT APPLICATION NUMBER: US/09/872,733
;  CURRENT FILING DATE:  2001-06-01
;  PRIOR APPLICATION NUMBER: PCT/US00/34985
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/173,036
;  PRIOR FILING DATE: 1999-12-23
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 2467
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  Mutated Human
;   OTHER INFORMATION: Immunodeficiency Virus - 1 Pol gene
US-09-872-733-3

  Query Match             32.9%;  Score 843;  DB 10;  Length 2467;
  Best Local Similarity   59.7%;  Pred. No. 3e-193;
  Matches 1435;  Conservative    0;  Mismatches  965;  Indels    3;  Gaps    1;

Qy    115 TGTGAAAAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAAC 174

Db      1 TGTACAGAGATGGAGAAGGAAGGGAAGATCAGCAAGATCGGGCCTGAGAACCCCTACAAC 60

Qy    175 ACCCCCACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTT 234

Db     61 ACTCCAGTCTTCGCAATCAAGAAGAAGGACAGTACCAAGTGGAGAAAGCTGGTGGACTTC 120

Qy    235 AGGGAACTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCT 294

Db    121 AGAGAGCTGAACAAGAGAACTCAGGACTTCTGGGAAGTTCAGCTGGGCATCCCACATCCC 180

Qy    295 GCAGGACTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCC 354

Db    181 GCTGGGTTGAAGAAGAAGAAGTCAGTGACAGTGCTGGATGTGGGTGATGCCTACTTCTCC 240

Qy    355 ATACCTCTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAAT 414

Db    241 GTTCCCTTGGACGAGGACTTCAGGAAGTACACTGCCTTCACGATACCTAGCATCAACAAC 300

Qy    415 GCAGAGCCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCA 474

Db    301 GAGACACCAGGCATCCGCTACCAGTACAACGTGCTGCCACAGGGATGGAAGGGATCACCA 360

Qy    475 GCCATCTTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGAT 534

Db    361 GCCATCTTTCAAAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAAAACCCAGAC 420

Qy    535 GTGACCTTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAA 594

Db    421 ATCGTGATCTATCAGTACATGGACGACCTCTACGTAGGAAGTGACCTGGAGATCGGGCAG 480

Qy    595 CATGACAGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCA 654

Db    481 CACAGGACCAAGATCGAGGAGCTGAGACAGCATCTGTTGAGGTGGGGACTGACCACACCA 540

Qy    655 GAAGAGAAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACA 714

Db    541 GACAAGAAGCACCAGAAGGAACCTCCCTTCCTGTGGATGGGCTACGAACTGCATCCTGAC 600

Qy    715 AAATGGAAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATA 774

Db    601 AAGTGGACAGTGCAGCCCATCGTGCTGCCTGAGAAGGACAGCTGGACTGTGAACGACATA 660

Qy    775 CAGAAGTTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAA 834

Db    661 CAGAAGCTCGTGGGCAAGTTGAACTGGGCAAGCCAGATCTACCCAGGCATCAAAGTTAGG 720

Qy    835 CATCTCTGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAG 894

Db    721 CAGCTGTGCAAGCTGCTTCGAGGAACCAAGGCACTGACAGAAGTGATCCCACTGACAGAG 780

Qy    895 ATGGCAGAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGT 954

Db    781 GAAGCAGAGCTAGAACTGGCAGAGAACCGAGAGATCCTGAAGGAGCCAGTACATGGAGTG 840

Qy    955 TATTACCAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGG 1014

Db    841 TACTACGACCCAAGCAAGGACCTGATCGCAGAGATCCAGAAGCAGGGGCAAGGCCAATGG 900

Qy   1015 TCTTATAAAATTCACCAAGA---AGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATA 1071

Db    901 ACCTACCAAATCTACCAGGAGCCCTTCAAGAACCTGAAGACAGGCAAGTACGCAAGGATG 960

Qy   1072 AAGAATACACATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAG 1131

Db    961 AGGGGTGCCCACACCAACGATGTGAAGCAGCTGACAGAGGCAGTGCAGAAGATCACCACA 1020

Qy   1132 GAAGCAATAGTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTA 1191

Db   1021 GAGAGCATCGTGATCTGGGGCAAGACTCCCAAGTTCAAGCTGCCCATACAGAAGGAGACA 1080

Qy   1192 TGGGAACAGTGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATC 1251

Db   1081 TGGGAGACATGGTGGACCGAGTACTGGCAAGCCACCTGGATCCCTGAGTGGGAGTTCGTG 1140

Qy   1252 TCAACACCACCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAA 1311

Db   1141 AACACCCCTCCCTTGGTGAAACTGTGGTATCAGCTGGAGAAGGAACCCATCGTGGGAGCA 1200

Qy   1312 GAAACCTATTATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATAT 1371

Db   1201 GAGACCTTCTACGTGGATGGGGCAGCCAACAGGGAGACCAAGCTGGGCAAGGCAGGCTAC 1260

Qy   1372 ATCACAGATAGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCA 1431

Db   1261 GTGACCAACCGAGGACGACAGAAAGTGGTGACCCTGACTGACACCACCAACCAGAAGACT 1320

Qy   1432 GAATTGGAAGCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTA 1491

Db   1321 GAGCTGCAAGCCATCTACCTAGCTCTGCAAGACAGCGGACTGGAAGTGAACATCGTGACA 1380

Qy   1492 GATTCACAATATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTA 1551

Db   1381 GACTCACAGTACGCACTGGGCATCATCCAAGCACAACCAGACCAATCCGAGTCAGAGCTG 1440

Qy   1552 GTTAATCAAATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCA 1611

Db   1441 GTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAAGTGTACCTGGCATGGGTACCA 1500
```

```
Qy   1612 GCACACAAAGGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGA 1671
          |||||||||||| || ||||||||  || || ||||  |  |||| |||   |||||  |
Db   1501 GCACACAAAGGAATTGGAGGAAATGAACAAGTAGATAAATTAGTCAGTGCTGGGATCCGG 1560

Qy   1672 CAAGTTCTCTTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGT 1731
           | || || ||| ||||   ||| ||    || ||||| |||||||| || |||||
Db   1561 AAGGTGCTGTTCCTGGACGGGATCGATAAGGCCCAAGATGAACATGAGAAGTACCACTCC 1620

Qy   1732 AATGTAAAAGAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTA 1791
          ||       |  |||    | | ||     | ||     |||| || | | | ||||||
Db   1621 AACTGGCGCGCTATGGCCAGCGACTTCAACCTGCCACCTGTAGTAGCAAAAGAAATAGTA 1680

Qy   1792 GACACCTGTGATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGAT 1851
          | || ||||||||||||||| |  |||||||| || || ||||| || | | | |     |
Db   1681 GCCAGCTGTGATAAATGTCAGCTAAAAGGAGAAGCCATGCATGGACAAGTAGACTGTAGT 1740

Qy   1852 CTAGGGACTTGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTA 1911
          | ||| |  ||||| |||| || || || || ||||| ||  | ||| | || |||||
Db   1741 CCAGGAATATGGCAGCTGGACTGCACGCACCTGGAGGGGAAGGTGATCCTGGTAGCAGTT 1800

Qy   1912 CATGTAGCTAGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACA 1971
          |||||||| |||||||  |||||||||| || || ||    ||.|| ||     | |||
Db   1801 CATGTAGCCAGTGGATATATAGAAGCAGAAGTTATCCCTGCTGAAACTGGGCAGGAAACA 1860

Qy   1972 GCACTATTTCTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAAT 2031
          |||   ||||| |||||||| ||||| |||||||  | | |    |||||| || ||
Db   1861 GCATATTTTCTTTTAAAATTAGCAGGAAGATGGCCAGTAAAAACAATACACACGGACAAC 1920

Qy   2032 GGTGCTAACTTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCAC 2091
          ||    ||||| ||         || |||   |     |||||||| || ||  ||||
Db   1921 GGAAGCAACTTCACTGGTGCTACGGTTAAGGCCGCCTGTTGGTGGGCGGGAATCAAGCAG 1980

Qy   2092 ACCTTTGGGGTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCAC 2151
             |||||  | || |||||||| ||    || |||| ||||    |||||  |  |
Db   1981 GAATTTGGAATTCCCTACAATCCCCAATCGCAAGGAGTCGTGGAGAGCATGAACAAGGAG 2040

Qy   2152 CTGAAAAATCAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTA 2211
          ||||| ||    || |    | | ||||| || ||  |     |  | ||   |||
Db   2041 CTGAAGAAGATCATCGGACAAGTGAGGGATCAGGCTGAGCACCTGAAGACAGCAGTGCAG 2100

Qy   2212 ATGGCAGTTCATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCA 2271
          ||||||||      |   || || |||||||  || || || ||||      | | |||
Db   2101 ATGGCAGTGTTCATCCACAACTTCAAAAGAAAAGGGGGGATTGGGGGGTACAGTGCAGGG 2160

Qy   2272 GAAAGATTAATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAAC 2331
          |||||  |  | |||| ||| | || ||     | |  |    |  ||     | | |
Db   2161 GAAAGGATCGTGGACATCATCGCCACCGACATCCAAACCAAGGAGCTGCAGAAGCAGATC 2220

Qy   2332 TCAAAATTTAAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGA 2391
           | ||  |  | || || ||||| || ||| | || .|| |  | | ||||||||||||||
Db   2221 ACCAAGATCCAGAACTTCCGGGTGTACTACCGCGACAGCCGCAACCCACTGTGGAAGGGA 2280

Qy   2392 CCCGGTGAGCTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATT 2451
          || |   |||| | ||||| || || || |||||  |  |  |||      |  |||||
Db   2281 CCAGCAAAGCTCCTCTGGAAGGGAGAGGGGGCAGTGGTGATCCAGGACAACAGTGACATC 2340

Qy   2452 AAGGTAGTACCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTG 2511
          || || || ||  | |||| || || |||   || ||||||  |    |  |  |
Db   2341 AAAGTGGTGCCAAGGCGCAAGGCCAAGATCATCCGCGACTATGGAAAACAGATGGCAGGT 2400

Qy   2512 GAT 2514
          |||
Db   2401 GAT 2403


RESULT 12
US-09-999-183-2
; Sequence 2, Application US/09999183
; Patent No. US20020147169A1
; GENERAL INFORMATION:
;  APPLICANT: MITROPHANOUS, et al
;  TITLE OF INVENTION: In Vivo Selection Method
;  FILE REFERENCE: 674523-2009
;  CURRENT APPLICATION NUMBER: US/09/999,183
;  CURRENT FILING DATE:  2001-11-29
;  PRIOR APPLICATION NUMBER: PCT/GB00/02136
;  PRIOR FILING DATE: 2000-06-02
;  PRIOR APPLICATION NUMBER: 9912965.2
;  PRIOR FILING DATE: 1999-06-03
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: SeqWin99
; SEQ ID NO 2
;   LENGTH: 4307
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Codon optimised gagpol sequence
US-09-999-183-2

  Query Match             32.5%;  Score 832.2;  DB 10;  Length 4307;
  Best Local Similarity   58.4%;  Pred. No. 1.5e-190;
  Matches 1471;  Conservative    0;  Mismatches 1043;  Indels    3;  Gaps    1;

Qy      1 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 60
            || |||||         | ||| | |||   ||      | ||||| || | ||| || ||
Db   1758 TTCCCCATTAGCCCTATCGAGACGGTACCGGTGAAGCTGAAGCCCGGGATGGACGGCCCG 1817

Qy     61 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 120
           ||  | ||||| ||||||||  || | || |||||    ||| |    || || ||   |
Db   1818 AAGGTCAAGCAATGGCCATTGACAGAGGAGAAGATCAAGGCACTGGTGGAGATTTGCACA 1877

Qy    121 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 180
           |||||||||||||| ||  |  |     |   |   ||    || || |||||||| ||
Db   1878 GAGATGGAAAAGGAAGGGAAAATCTCCAAGATTGGGCCTGAGAACCCGTACAACACGCCG 1937

Qy    181 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 240
            || || || ||||| ||||   ||  |||||| | | |||| | || ||  | ||
Db   1938 GTGTTCGCAATCAAGAAGAAGGACTCGACGAAATGGCGCAAGCTGGTGGACTTCCGCGAG 1997

Qy    241 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 300
          || || | |  ||| || |||||   ||| || ||  | || || ||| ||||| |||||
Db   1998 CTGAACAAGCGCACGCAAGACTTCTGGGAGGTTCAGCTGGGCATCCCGCACCCCGCAGGG 2057

Qy    301 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 360
          ||    || | ||||    | || |||||||| | |||||||| || |||||| | ||
Db   2058 CTGAAGAAGAAGAAATCCGTGACCGTACTGGATGTGGGTGATGCCTACTTCTCCGTTCCC 2117

Qy    361 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 420
          || || ||||| || ||| |||||||||||| ||  | || ||  | || || | ||
Db   2118 CTGGACGAAGACTTCAGGAAGTACACTGCCTTCACAATCCCTTCGATCAACAACGAGACA 2177

Qy    421 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 480
          || || |  |||||    || || || ||||| ||||| |||| || || || || |||
Db   2178 CCGGGGATTCGATATCAGTACAACGTGCTGCCCCAGGGCTGGAAAGGCTCTCCCGCAATC 2237

Qy    481 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 540
          |||||    |  ||||   |  | || || || ||| | ||    || || || |  |
Db   2238 TTCCAGAGTAGCATGACCAAAATCCTGGAGCCTTTCCGCAAACAGAACCCCGACATCGTC 2297

Qy    541 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 600
          |    |||| |||||||| | |   | |   |||             | | ||   |
Db   2298 ATCTATCAGTACATGGATGACTTGTACGTGGGCTCTGATCTAGAGATAGGGCAGCACCGC 2357

Qy    601 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 660
          |      |    ||      | | | ||   ||   |  |  | || || ||  ||
Db   2358 ACCAAGATCGAGGAGCTGCGCCAGCACCTGTTGAGGTGGGGACTGACCACACCCGACAAG 2417

Qy    661 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 720
          ||   ||| || || || || || | |||||||| ||||  ||   ||    ||||||
Db   2418 AAGCACCAGAAGGAGCCTCCCTTCCTCTGGATGGGTTACGAGCTGCACCCTGACAAATGG 2477

Qy    721 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 780
          |   ||||    || | | ||||| | | ||| | |||||| || || || |||||||||
Db   2478 ACCGTGCAGCCTATCGTGCTGCCAGAGAAAGACAGCTGGACTGTCAACGACATACAGAAG 2537

Qy    781 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 840
```

```
Db   2538 CTGGTGGGGAAGTTGAACTGGGCCAGTCAGATTTACCCAGGGATTAAGGTGAGGCAGCTG 2597

Qy    841 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 900

Db   2598 TGCAAACTCCTCCGCGGAACCAAGGCACTCACAGAGGTGATCCCCCTAACCGAGGAGGCC 2657

Qy    901 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 960

Db   2658 GAGCTCGAACTGGCAGAAAACCGAGAGATCCTAAAGGAGCCCGTGCACGGCGTGTACTAT 2717

Qy    961 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 1020

Db   2718 GACCCCTCCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGGCAAGGCCAGTGGACCTAT 2777

Qy   1021 AAAATTCACCAAGA---AGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAAT 1077

Db   2778 CAGATTTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCCGGATGAGGGGT 2837

Qy   1078 ACACATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCA 1137

Db   2838 GCCCACACTAACGACGTCAAGCAGCTGACCGAGGCCGTGCAGAAGATCACCACCGAAAGC 2897

Qy   1138 ATAGTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAA 1197

Db   2898 ATCGTGATCTGGGGAAAGACTCCTAAGTTCAAGCTGCCCATCCAGAAGGAAACCTGGGAA 2957

Qy   1198 CAGTGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACA 1257

Db   2958 ACCTGGTGGACAGAGTATTGGCAGGCCACCTGGATTCCTGAGTGGGAGTTCGTCAACACC 3017

Qy   1258 CCACCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACC 1317

Db   3018 CCTCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATAGTGGGCGCCGAAACC 3077

Qy   1318 TATTATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACA 1377

Db   3078 TTCTACGTGGATGGGGCCGCTAACAGGGAGACTAAGCTGGGCAAAGCCGGATACGTCACT 3137

Qy   1378 GATAGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTG 1437

Db   3138 AACCGGGGCAGACAGAAGGTTGTCACCCTCACTGACACCACCAACCAGAAGACTGAGCTG 3197

Qy   1438 GAAGCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCA 1497

Db   3198 CAGGCCATTTACCTCGCTTTGCAGGACTCGGGCCTGGAGGTGAACATCGTGACAGACTCT 3257

Qy   1498 CAATATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAAT 1557

Db   3258 CAGTATGCCCTGGGCATCATTCAAGCCCAGCCAGACCAGAGTGAGTCCGAGCTGGTCAAT 3317

Qy   1558 CAAATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACAC 1617

Db   3318 CAGATCATCGAGCAGCTGATCAAGAAGGAAAAGGTCTATCTGGCCTGGGTACCCGCCCAC 3377

Qy   1618 AAAGGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTT 1677

Db   3378 AAAGGCATTGGCGGCAATGAGCAGGTCGACAAGCTGGTCTCGGCTGGCATCAGGAAGGTG 3437

Qy   1678 CTCTTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTA 1737

Db   3438 CTATTCCTGGATGGCATCGACAAGGCCCAGGACGAGCACGAGAAATACCACAGCAACTGG 3497

Qy   1738 AAAGAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACC 1797

Db   3498 CGGGCCATGGCTAGCGACTTCAACCTGCCCCCTGTGGTGGCCAAAGAGATCGTGGCCAGC 3557

Qy   1798 TGTGATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATTCAGATCTAGGG 1857

Db   3558 TGTGACAAGTGTCAGCTCAAGGGCGAAGCCATGCATGGCCAGGTGGACTGTAGCCCCGGC 3617

Qy   1858 ACTTGGCAAATGGATTGTACCCATCTAGAGGGAAAAATAATCATAGTTGCAGTACATGTA 1917

Db   3618 ATCTGGCAACTCGATTGCACCCATCTGGAGGGCAAGGTTATCCTGGTAGCCGTCCATGTG 3677

Qy   1918 GCTAGTGGATTCATAGAAGCAGAGGTAATTCCACAAGAGACAGGAAGACAGACAGCACTA 1977

Db   3678 GCCAGTGGCTACATCGAGGCCGAGGTCATTCCCGCCGAAACAGGGCAGGAGACAGCCTAC 3737

Qy   1978 TTTCTGTTAAAATTGGCAGGCAGATGGCCTATTACACATCTACACACAGATAATGGTGCT 2037

Db   3738 TTCCTCCTGAAGCTGGCAGGCCGGTGGCCAGTGAAGACCATCCATACTGACAATGGCAGC 3797

Qy   2038 AACTTTGCTTCGCAAGAAGTAAAGATGGTTGCATGGTGGGCAGGGATAGAGCACACCTTT 2097

Db   3798 AATTTCACCAGTGCTACGGTTAAGGCCGCCTGCTGGTGGGCGGGAATCAAGCAGGAGTTC 3857

Qy   2098 GGGGTACCATACAATCCACAGAGTCAGGGAGTAGTGGAAGCAATGAATCACCACCTGAAA 2157

Db   3858 GGGATCCCCTACAATCCCCAGAGTCAGGGCGTCGTCGAGTCTATGAATAAGGAGTTAAAG 3917

Qy   2158 AATCAAATAGATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCA 2217

Db   3918 AAGATTATCGGCCAGGTCAGAGATCAGGCTGAGCATCTCAAGACCGCGGTCCAAATGGCG 3977

Qy   2218 GTTCATTGCATGAATTTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGA 2277

Db   3978 GTATTCATCCACAATTTCAAGCGGAAGGGGGGGATTGGGGGGTACAGTGCGGGGGAGCGG 4037

Qy   2278 TTAATTAACATGATCACTACAGAACAAGAGATACAATTTCAACAATCAAAAAACTCAAAA 2337

Db   4038 ATCGTGGACATCATCGCGACCGACATCCAGACTAAGGAGCTGCAAAAGCAGATTACCAAG 4097

Qy   2338 TTTAAAAATTTTCGGGTCTATTACAGAGAAGGCAGAGATCAACTGTGGAAGGGACCCGGT 2397

Db   4098 ATTCAGAATTTCCGGGTCTACTACAGGGACAGCAGAAATCCCCTCTGGAAAGGCCCAGCG 4157

Qy   2398 GAGCTATTGTGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGGACAGACATTAAGGTA 2457

Db   4158 AAGCTCCTCTGGAAGGGTGAGGGGCAGTAGTGATCCAGGATAATAGCGACATCAAGGTG 4217

Qy   2458 GTACCCAGAAGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAAGAGGTGGAT 2514

Db   4218 GTGCCCAGAAGAAAGGCGAAGATCATTAGGGATTATGGCAAACAGATGGCGGGTGAT 4274


RESULT 13
US-09-735-487-7
; Sequence 7, Application US/09735487
; Patent No. US20020042679A1
; GENERAL INFORMATION:
;  APPLICANT:  de BETHUNE, Marie-Pierre
;  APPLICANT:  HERTOGS, Kurt
;  APPLICANT:  PAUWELS, Rudi
;  TITLE OF INVENTION:  METHOD OF MANAGING THE CHEMOTHERAPY OF PATIENTS WHO ARE
;  TITLE OF INVENTION:  HIV POSTITIVE BASED ON THE PHENOTYPIC DRUG SENSITIVITY
;  TITLE OF INVENTION:  OF HUMAN HIV STRAINS
;  FILE REFERENCE:  1377-125P
;  CURRENT APPLICATION NUMBER: US/09/735,487
;  CURRENT FILING DATE:  2000-12-14
;  PRIOR APPLICATION NUMBER:  09/117,217
;  PRIOR FILING DATE:  1998-07-24
;  NUMBER OF SEQ ID NOS:  15
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 7
;   LENGTH: 2601
;   TYPE: DNA
;   ORGANISM: HIV-HXB2
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(492)
;   OTHER INFORMATION: gag Polyprotein
US-09-735-487-7

  Query Match             30.2%;  Score 775.4;  DB 10;  Length 2601;
  Best Local Similarity   64.1%;  Pred. No. 5.6e-177;
```

```
  Matches 1185;  Conservative    0;  Mismatches  661;  Indels    3;  Gaps    1;

Qy      1 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 60
          ||||||||         | ||| |||||  |||    ||||||||||||| ||||||  |||
Db    747 TTTCCCATTAGCCCTATTGAGACTGTACCAGTAAAATTAAAGCCAGGAATGGATGGCCCA 806

Qy     61 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 120
          ||| | || || ||||||||  || |||||| |||   ||||||  |||||| ||| |
Db    807 AAAGTTAAACAATGGCCATTGACAGAAGAAAAAATAAAAGCATTAGTAGAAATTTGTACA 866

Qy    121 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 180
          |||||||||||||| |   |  |     ||  |   ||    ||||||||||| || ||
Db    867 GAGATGGAAAAGGAAGCGAAAATTTCAAAAATTGGGCCTGAAAATCCATACAATACTCCA 926

Qy    181 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 240
          ||||||  |||||||||||  || |  |  |||||||||   |  |||||||  || |||
Db    927 GTATTTGCCATAAAGAAAAAAGACAGTACTAAATGGAGAAAATTAGTAGATTTCAGAGAA 986

Qy    241 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 300
          || |||| |   ||||| |||||  |||||| ||||||||||||||||| || |||||
Db    987 CTTAATAAGAGAACTCAAGACTTCTGGGAAGTTCAATTAGGAATACCACATCCCGCAGGC 1046

Qy    301 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 360
          ||  ||| |  |||  | | |||||||||||| | ||||||||||||||| ||  | ||
Db   1047 TTAAAAAAGAAAAAATCAGTAACAGTACTGGATGTGGGTGATGCATATTTTTCAGTTCCC 1106

Qy    361 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 420
          ||||||||||  || ||| |||| |||||  ||||    |||| |||||
Db   1107 TTAGATGAAGACTTCAGGAAGTATACTGCATTTACCATACCTAGTATAAACAATGAGACA 1166

Qy    421 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 480
          |||||  |   ||||    || || || || || ||||||||||  || ||||||||  ||
Db   1167 CCAGGGATTAGATATCAGTACAATGTGCTTCCACAGGGATGGAAAGGATCACCAGCAATA 1226

Qy    481 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 540
          ||||||   |  |||| | |  |  |||| || || || ||  |||||||||  |
Db   1227 TTCCAAAGTAGCATGACAAAAATCTTAGAGCCTTTTAGAAAACAAAATCCAGACATAGTT 1286

Qy    541 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 600
          |   || || |||||||| | |  |||   ||||      |    | | |||
Db   1287 ATCTATCAATACATGGATGATTTGTATGTAGGATCTGACTTAGAAATAGGGCAGCATAGA 1346

Qy    601 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 660
          |   | |    ||   |  |||  ||  ||   ||  || | || |||||  |
Db   1347 ACAAAAATAGAGGAGCTGAGACAACATCTGTTGAGGTGGGGACTTACCACACCAGACAAA 1406

Qy    661 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 720
          |||   || ||||| || ||||| |  |||||||| || ||| |    ||    ||||||
Db   1407 AAACATCAGAAAGAACCTCCATTCCTTTGGATGGGTTATGAACTCCATCCTGATAAATGG 1466

Qy    721 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 780
          |   | ||    |||| | ||||| ||| ||| | |||||| || ||||| ||||||||
Db   1467 ACAGTACAGCCTATAGTGCTGCCAGAAAAAGACAGCTGGACTGTCAATGACATACAGAAG 1526

Qy    781 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 840
          ||||| ||   ||| |||||||||  ||| ||||| ||||| || |||   |  ||  |
Db   1527 TTAGTGGGGAAATTGAATTGGGCAAGTCAGATTTACCCAGGGATTAAAGTAAGGCAATTA 1586

Qy    841 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 900
          ||||   |  |||||||||  |   | |||||||| | | | |     || ||    |||
Db   1587 TGTAAACTCCTTAGAGGAACCAAAGCACTAACAGAAGTAATACCACTAACAGAAGAAGCA 1646

Qy    901 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 960
          ||   ||||   |  ||||| | |   ||||| |   |   | | ||| |||||
Db   1647 GAGCTAGAACTGGCAGAAAACAGAGAGATTCTAAAAGAACCAGTACATGGAGTGTATTAT 1706

Qy    961 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 1020
          |   |   ||    |||  |||     ||   |||     | |   || ||| | |||
Db   1707 GACCCATCAAAAGACTTAATAGCAGAAATACAGAAGCAGGGGCAAGGCCAATGGACATAT 1766

Qy    1021 AAAATTCACCAAGA---AGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAAT 1077
           ||||| | |||||   |  ||||  ||||||  |||||||| |||||  || | |  |
Db    1767 CAAATTTATCAAGAGCCATTTAAAAATCTGAAAACAGGAAAATATGCAAGAATGAGGGGT 1826

Qy    1078 ACACATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCA 1137
           | || || ||||  || | || |||| || | | | | || ||||||   |  |||
Db    1827 GCCCACACTAATGATGTAAAACAATTAACAGAGGCAGTGCAAAAAATAACCACAGAAAGC 1886

Qy    1138 ATAGTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAA 1197
           ||||| || |||||| ||   || |||||  |  | ||  |  | |||||   |||||||
Db    1887 ATAGTAATATGGGGAAAGACTCCTAAATTTAAACTGCCCATACAAAAGGAAACATGGGAA 1946

Qy    1198 CAGTGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACA 1257
           ||||||||||| |||||||| | ||||||| || || ||||| ||| |    ||
Db    1947 ACATGGTGGACAGAGTATTGGCAAGCCACCTGGATTCCTGAGTGGGAGTTTGTTAATACC 2006

Qy    1258 CCACCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACC 1317
           || ||  |||| | ||||   | | |  ||| ||| || || ||||  |||| |||||||
Db    2007 CCTCCCTTAGTGAAATTATGGTACCAGTTAGAGAAAGAACCCATAGTAGGAGCAGAAACC 2066

Qy    1318 TATTATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACA 1377
           |  |||  ||||||  ||  ||| |   || | ||| ||| |||||||||||| | ||
Db    2067 TTCTATGTAGATGGGGCAGCTAACAGGGAGACTAAATTAGGAAAAGCAGGATATGTTACT 2126

Qy    1378 GATAGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTG 1437
           ||||  || | | | |||||        ||    | || || |||||  |  | || ||
Db    2127 AATAGAGGAAGACAAAAAGTTGTCACCCTAACTGACACAACAAATCAGAAGACTGAGTTA 2186

Qy    1438 GAAGCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCA 1497
           |||||  ||    | || |||   || || ||   | | | ||| ||  ||  ||| |||
Db    2187 CAAGCAATTTATCTAGCTTTGCAGGATTCGGGATTAGAAGTAAACATAGTAACAGACTCA 2246

Qy    1498 CAATATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAAT 1557
           |||||||   | ||||| ||   || | ||   ||    ||      | |||| |||
Db    2247 CAATATGCATTAGGAATCATTCAAGCACAACCAGATCAAAGTGAATCAGAGTTAGTCAAT 2306

Qy    1558 CAAATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACAC 1617
           ||||||||||  |  | || ||||||  | |  | ||| | ||||||||||||||||||
Db    2307 CAAATAATAGAGCAGTTAATAAAAAAGGAAAAGGTCTATCTGGCATGGGTACCAGCACAC 2366

Qy    1618 AAAGGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTT 1677
           |||||  || |||||||  || || ||||  |  |||| |||   || || ||  ||||
Db    2367 AAAGGAATTGGAGGAAATGAACAAGTAGATAAATTAGTCAGTGCTGGAATCAGGAAAGTA 2426

Qy    1678 CTCTTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTA 1737
           || || || ||    |||||    || ||||| |||||||| ||||| || ||||||
Db    2427 CTATTTTTAGATGGAATAGATAAGGCCCAAGATGAACATGAGAAATATCACAGTAATTGG 2486

Qy    1738 AAAGAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACC 1797
           | || | |||      | |||    | ||     |||| || | | | ||||||| || |
Db    2487 AGAGCAATGGCTAGTGATTTTAACCTGCCACCTGTAGTAGCAAAAGAAATAGTAGCCAGC 2546

Qy    1798 TGTGATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATT 1846
           |||||||||||||| |  |||||||| || || ||||| || | | | |
Db    2547 TGTGATAAATGTCAGCTAAAAGGAGAAGCCATGCATGGACAAGTAGACT 2595


RESULT 14
US-09-735-487-9
; Sequence 9, Application US/09735487
; Patent No. US20020042679A1
; GENERAL INFORMATION:
;  APPLICANT:  de BETHUNE, Marie-Pierre
;  APPLICANT:  HERTOGS, Kurt
;  APPLICANT:  PAUWELS, Rudi
;  TITLE OF INVENTION:  METHOD OF MANAGING THE CHEMOTHERAPY OF PATIENTS WHO ARE
;  TITLE OF INVENTION:  HIV POSTITIVE BASED ON THE PHENOTYPIC DRUG SENSITIVITY
;  TITLE OF INVENTION:  OF HUMAN HIV STRAINS
;  FILE REFERENCE:  1377-125P
;  CURRENT APPLICATION NUMBER: US/09/735,487
;  CURRENT FILING DATE:  2000-12-14
;  PRIOR APPLICATION NUMBER:  09/117,217
;  PRIOR FILING DATE:  1998-07-24
;  NUMBER OF SEQ ID NOS:  15
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 9
```

```
;   LENGTH: 2601
;   TYPE: DNA
;   ORGANISM: HIV-HXB2
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (334)..(489)
;   OTHER INFORMATION: gag P6 (52 AA)
US-09-735-487-9

  Query Match           30.2%;  Score 775.4;  DB 10;  Length 2601;
  Best Local Similarity  64.1%;  Pred. No. 5.6e-177;
  Matches 1185;  Conservative    0;  Mismatches  661;  Indels    3;  Gaps    1;

Qy       1 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 60
Db     747 TTTCCCATTAGCCCTATTGAGACTGTACCAGTAAAATTAAAGCCAGGAATGGATGGCCCA 806

Qy      61 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 120
Db     807 AAAGTTAAACAATGGCCATTGACAGAAGAAAAAATAAAAGCATTAGTAGAAATTTGTACA 866

Qy     121 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 180
Db     867 GAGATGGAAAAGGAAGCGAAAATTTCAAAAATTGGGCCTGAAAATCCATACAATACTCCA 926

Qy     181 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 240
Db     927 GTATTTGCCATAAAGAAAAAAGACAGTACTAAATGGAGAAAATTAGTAGATTTCAGAGAA 986

Qy     241 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 300
Db     987 CTTAATAAGAGAACTCAAGACTTCTGGGAAGTTCAATTAGGAATACCACATCCCGCAGGC 1046

Qy     301 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 360
Db    1047 TTAAAAAAGAAAAAATCAGTAACAGTACTGGATGTGGGTGATGCATATTTTTCAGTTCCC 1106

Qy     361 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 420
Db    1107 TTAGATGAAGACTTCAGGAAGTATACTGCATTTACCATACCTAGTATAAACAATGAGACA 1166

Qy     421 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 480
Db    1167 CCAGGGATTAGATATCAGTACAATGTGCTTCCACAGGGATGGAAAGGATCACCAGCAATA 1226

Qy     481 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 540
Db    1227 TTCCAAAGTAGCATGACAAAAATCTTAGAGCCTTTTAGAAAACAAAATCCAGACATAGTT 1286

Qy     541 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 600
Db    1287 ATCTATCAATACATGGATGATTTGTATGTAGGATCTGACTTAGAAATAGGGCAGCATAGA 1346

Qy     601 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 660
Db    1347 ACAAAAATAGAGGAGCTGAGACAACATCTGTTGAGGTGGGGACTTACCACACCAGACAAA 1406

Qy     661 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 720
Db    1407 AAACATCAGAAAGAACCTCCATTCCTTTGGATGGGTTATGAACTCCATCCTGATAAATGG 1466

Qy     721 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 780
Db    1467 ACAGTACAGCCTATAGTGCTGCCAGAAAAAGACAGCTGGACTGTCAATGACATACAGAAG 1526

Qy     781 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 840
Db    1527 TTAGTGGGGAAATTGAATTGGGCAAGTCAGATTTACCCAGGGATTAAAGTAAGGCAATTA 1586

Qy     841 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 900
Db    1587 TGTAAACTCCTTAGAGGAACCAAAGCACTAACAGAAGTAATACCACTAACAGAAGAAGCA 1646

Qy     901 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 960
Db    1647 GAGCTAGAACTGGCAGAAAACAGAGAGATTCTAAAAGAACCAGTACATGGAGTGTATTAT 1706

Qy     961 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 1020
Db    1707 GACCCATCAAAAGACTTAATAGCAGAAATACAGAAGCAGGGGCAAGGCCAATGGACATAT 1766

Qy    1021 AAAATTCACCAAGA---AGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAAT 1077
Db    1767 CAAATTTATCAAGAGCCATTTAAAAATCTGAAAACAGGAAAATATGCAAGAATGAGGGGT 1826

Qy    1078 ACACATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCA 1137
Db    1827 GCCCACACTAATGATGTAAAACAATTAACAGAGGCAGTGCAAAAAATAACCACAGAAAGC 1886

Qy    1138 ATAGTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAA 1197
Db    1887 ATAGTAATATGGGGAAAGACTCCTAAATTTAAACTGCCCATACAAAAGGAAACATGGGAA 1946

Qy    1198 CAGTGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACA 1257
Db    1947 ACATGGTGGACAGAGTATTGGCAAGCCACCTGGATTCCTGAGTGGGAGTTTGTTAATACC 2006

Qy    1258 CCACCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACC 1317
Db    2007 CCTCCCTTAGTGAAATTATGGTACCAGTTAGAGAAAGAACCCATAGTAGGAGCAGAAACC 2066

Qy    1318 TATTATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACA 1377
Db    2067 TTCTATGTAGATGGGGCAGCTAACAGGGAGACTAAATTAGGAAAAGCAGGATATGTTACT 2126

Qy    1378 GATAGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTG 1437
Db    2127 AATAGAGGAAGACAAAAAGTTGTCACCCTAACTGACACAACAAATCAGAAGACTGAGTTA 2186

Qy    1438 GAAGCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCA 1497
Db    2187 CAAGCAATTTATCTAGCTTTGCAGGATTCGGGATTAGAAGTAAACATAGTAACAGACTCA 2246

Qy    1498 CAATATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAAT 1557
Db    2247 CAATATGCATTAGGAATCATTCAAGCACAACCAGATCAAAGTGAATCAGAGTTAGTCAAT 2306

Qy    1558 CAAATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACAC 1617
Db    2307 CAAATAATAGAGCAGTTAATAAAAAAGGAAAAGGTCTATCTGGCATGGGTACCAGCACAC 2366

Qy    1618 AAAGGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTT 1677
Db    2367 AAAGGAATTGGAGGAAATGAACAAGTAGATAAATTAGTCAGTGCTGGAATCAGGAAAGTA 2426

Qy    1678 CTCTTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTA 1737
Db    2427 CTATTTTTAGATGGAATAGATAAGGCCCAAGATGAACATGAGAAATATCACAGTAATTGG 2486

Qy    1738 AAAGAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACC 1797
Db    2487 AGAGCAATGGCTAGTGATTTTAACCTGCCACCTGTAGTAGCAAAAGAAATAGTAGCCAGC 2546

Qy    1798 TGTGATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATT 1846
Db    2547 TGTGATAAATGTCAGCTAAAAGGAGAAGCCATGCATGGACAAGTAGACT 2595


RESULT 15
US-09-735-487-11
; Sequence 11, Application US/09735487
; Patent No. US20020042679A1
; GENERAL INFORMATION:
;  APPLICANT:  de BETHUNE, Marie-Pierre
;  APPLICANT:  HERTOGS, Kurt
;  APPLICANT:  PAUWELS, Rudi
```

```
;  TITLE OF INVENTION:  METHOD OF MANAGING THE CHEMOTHERAPY OF PATIENTS WHO ARE
;  TITLE OF INVENTION:  HIV POSTITIVE BASED ON THE PHENOTYPIC DRUG SENSITIVITY
;  TITLE OF INVENTION:  OF HUMAN HIV STRAINS
;  FILE REFERENCE:  1377-125P
;  CURRENT APPLICATION NUMBER: US/09/735,487
;  CURRENT FILING DATE:  2000-12-14
;  PRIOR APPLICATION NUMBER:  09/117,217
;  PRIOR FILING DATE:  1998-07-24
;  NUMBER OF SEQ ID NOS:  15
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 11
;   LENGTH: 2601
;   TYPE: DNA
;   ORGANISM: HIV-HXB2
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (453)..(749)
;   OTHER INFORMATION: Protease
US-09-735-487-11

  Query Match             30.2%;  Score 775.4;  DB 10;  Length 2601;
  Best Local Similarity   64.1%;  Pred. No. 5.6e-177;
  Matches 1185;  Conservative    0;  Mismatches  661;  Indels    3;  Gaps    1;

Qy       1 TTTCCCATAGCTAAAGTAGAGCCTGTAAAAGTCGCCTTAAAGCCAGGAAAGGATGGACCA 60
Db     747 TTTCCCATTAGCCCTATTGAGACTGTACCAGTAAAATTAAAGCCAGGAATGGATGGCCCA 806

Qy      61 AAATTGAAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAA 120
Db     807 AAAGTTAAACAATGGCCATTGACAGAAGAAAAAATAAAAGCATTAGTAGAAATTTGTACA 866

Qy     121 AAGATGGAAAAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCC 180
Db     867 GAGATGGAAAAGGAAGCGAAAATTTCAAAAATTGGGCCTGAAAATCCATACAATACTCCA 926

Qy     181 ACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAA 240
Db     927 GTATTTGCCATAAAGAAAAAAGACAGTACTAAATGGAGAAAATTAGTAGATTTCAGAGAA 986

Qy     241 CTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGA 300
Db     987 CTTAATAAGAGAACTCAAGACTTCTGGGAAGTTCAATTAGGAATACCACATCCCGCAGGC 1046

Qy     301 CTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCT 360
Db    1047 TTAAAAAAGAAAAAATCAGTAACAGTACTGGATGTGGGTGATGCATATTTTTCAGTTCCC 1106

Qy     361 CTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAG 420
Db    1107 TTAGATGAAGACTTCAGGAAGTATACTGCATTTACCATACCTAGTATAAACAATGAGACA 1166

Qy     421 CCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATC 480
Db    1167 CCAGGGATTAGATATCAGTACAATGTGCTTCCACAGGGATGGAAAGGATCACCAGCAATA 1226

Qy     481 TTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACC 540
Db    1227 TTCCAAAGTAGCATGACAAAAATCTTAGAGCCTTTTAGAAAACAAAATCCAGACATAGTT 1286

Qy     541 TTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGAC 600
Db    1287 ATCTATCAATACATGGATGATTTGTATGTAGGATCTGACTTAGAAATAGGGCAGCATAGA 1346

Qy     601 AGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAG 660
Db    1347 ACAAAAATAGAGGAGCTGAGACAACATCTGTTGAGGTGGGGACTTACCACACCAGACAAA 1406

Qy     661 AAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGG 720
Db    1407 AAACATCAGAAAGAACCTCCATTCCTTTGGATGGGTTATGAACTCCATCCTGATAAATGG 1466

Qy     721 AAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAG 780

Db    1467 ACAGTACAGCCTATAGTGCTGCCAGAAAAAGACAGCTGGACTGTCAATGACATACAGAAG 1526

Qy     781 TTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAAACATCTC 840
Db    1527 TTAGTGGGGAAATTGAATTGGGCAAGTCAGATTTACCCAGGGATTAAAGTAAGGCAATTA 1586

Qy     841 TGTAGGTTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCAGTGGACTGAGATGGCA 900
Db    1587 TGTAAACTCCTTAGAGGAACCAAAGCACTAACAGAAGTAATACCACTAACAGAAGAAGCA 1646

Qy     901 GAAGCAGAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTAC 960
Db    1647 GAGCTAGAACTGGCAGAAAACAGAGAGATTCTAAAAGAACCAGTACATGGAGTGTATTAT 1706

Qy     961 CAAGAAGGCAAGCCATTAGAAGCCACGGTAATAAAGAGTCAGGACAATCAGTGGTCTTAT 1020
Db    1707 GACCCATCAAAAGACTTAATAGCAGAAATACAGAAGCAGGGGCAAGGCCAATGGACATAT 1766

Qy    1021 AAAATTCACCAAGA---AGACAAAATACTGAAAGTAGGAAAATTTGCAAAGATAAAGAAT 1077
Db    1767 CAAATTTATCAAGAGCCATTTAAAAATCTGAAAACAGGAAAATATGCAAGAATGAGGGGT 1826

Qy    1078 ACACATACCAATGGAGTGAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCA 1137
Db    1827 GCCCACACTAATGATGTAAAACAATTAACAGAGGCAGTGCAAAAAATAACCACAGAAAGC 1886

Qy    1138 ATAGTGATCTGGGGACAGGTCCCAAAATTCCACTTACCAGTTGAGAAGGATGTATGGGAA 1197
Db    1887 ATAGTAATATGGGGAAAGACTCCTAAATTTAAACTGCCCATACAAAAGGAAACATGGGAA 1946

Qy    1198 CAGTGGTGGACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTATCTCAACA 1257
Db    1947 ACATGGTGGACAGAGTATTGGCAAGCCACCTGGATTCCTGAGTGGGAGTTTGTTAATACC 2006

Qy    1258 CCACCGCTAGTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACC 1317
Db    2007 CCTCCCTTAGTGAAATTATGGTACCAGTTAGAGAAAGAACCCATAGTAGGAGCAGAAACC 2066

Qy    1318 TATTATACAGATGGATCATGTAATAAACAGTCAAAAGAAGGGAAAGCAGGATATATCACA 1377
Db    2067 TTCTATGTAGATGGGGCAGCTAACAGGGAGACTAAATTAGGAAAAGCAGGATATGTTACT 2126

Qy    1378 GATAGGGGCAAAGACAAAGTAAAAGTGTTAGAACAGACTACTAATCAACAAGCAGAATTG 1437
Db    2127 AATAGAGGAAGACAAAAAGTTGTCACCCTAACTGACACAACAAATCAGAAGACTGAGTTA 2186

Qy    1438 GAAGCATTTCTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGATTCA 1497
Db    2187 CAAGCAATTTATCTAGCTTTGCAGGATTCGGGATTAGAAGTAAACATAGTAACAGACTCA 2246

Qy    1498 CAATATGTTATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAAT 1557
Db    2247 CAATATGCATTAGGAATCATTCAAGCACAACCAGATCAAAGTGAATCAGAGTTAGTCAAT 2306

Qy    1558 CAAATAATAGAAGAAATGATTAAAAAGTCAGAAATTTATGTAGCATGGGTACCAGCACAC 1617
Db    2307 CAAATAATAGAGCAGTTAATAAAAAAGGAAAAGGTCTATCTGGCATGGGTACCAGCACAC 2366

Qy    1618 AAAGGTATAGGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTT 1677
Db    2367 AAAGGAATTGGAGGAAATGAACAAGTAGATAAATTAGTCAGTGCTGGAATCAGGAAAGTA 2426

Qy    1678 CTCTTCTTGGAAAAGATAGAGCCAGCACAAGAAGAACATGATAAATACCATAGTAATGTA 1737
Db    2427 CTATTTTTAGATGGAATAGATAAGGCCCAAGATGAACATGAGAAATATCACAGTAATTGG 2486

Qy    1738 AAAGAATTGGTATTCAAATTTGGATTACCCAGAATAGTGGCCAGACAGATAGTAGACACC 1797
Db    2487 AGAGCAATGGCTAGTGATTTTAACCTGCCACCTGTAGTAGCAAAAGAAATAGTAGCCAGC 2546

Qy    1798 TGTGATAAATGTCATCAGAAAGGAGAGGCTATACATGGGCAGGCAAATT 1846
```

Db 2547 TGTGATAAATGTCAGCTAAAAGGAGAAGCCATGCATGGACAAGTAGACT 2595

Search completed: November 16, 2002, 13:12:54
Job time : 222.199 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 16, 2002, 07:32:18 ; Search time 2375.97 Seconds
(without alignments)
17477.179 Million cell updates/sec

Title: M33262_COPY_3105_5668
Perfect score: 2564
Sequence: 1 TTTCCCATAGCTAAAGTAGA.....GCTAGAGAGGTGGCATAGCC 2564

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	113.8	4.4	705	17	AQ475349	AQ475349 CITBI-E1-
	2	110.6	4.3	732	17	AQ422201	AQ422201 RPCI-11-1
	3	100.6	3.9	659	17	AZ853584	AZ853584 2M0156B20
c	4	100.4	3.9	604	17	AZ953622	AZ953622 2M0218O20
	5	98.2	3.8	610	17	BH295203	BH295203 CH230-104
	6	97.2	3.8	706	17	AZ362320	AZ362320 1M0107E20
	7	95.4	3.7	548	17	AQ008834	AQ008834 RPCI11-20
c	8	95.2	3.7	641	17	AZ720063	AZ720063 RPCI-24-1
c	9	94.8	3.7	704	17	AZ589335	AZ589335 1M0398D17
c	10	93.4	3.6	586	9	AI633818	AI633818 tt28h04.x
	11	93.2	3.6	620	17	AZ430794	AZ430794 1M0215K23
	12	93.2	3.6	629	17	AQ582156	AQ582156 RPCI-11-4
	13	93	3.6	480	17	AQ401360	AQ401360 HS_5064_A
c	14	92.4	3.6	708	17	AZ345218	AZ345218 1M0079H11
	15	92	3.6	511	17	BH287730	BH287730 CH230-181
c	16	91.6	3.6	647	17	AZ370928	AZ370928 1M0122G14
	17	91	3.5	560	17	AQ318206	AQ318206 RPCI11-11
c	18	90.8	3.5	805	17	AZ722830	AZ722830 RPCI-24-6
	19	89.2	3.5	439	17	AQ471959	AQ471959 CITBI-E1-
c	20	88.6	3.5	692	17	AZ937169	AZ937169 2M0195E16
c	21	88.2	3.4	421	17	AZ438768	AZ438768 1M0229F05
	22	88.2	3.4	514	17	AQ430887	AQ430887 HS_5137_A
	23	88	3.4	718	17	BH305009	BH305009 CH230-25L
c	24	87.8	3.4	713	17	BH062296	BH062296 RPCI-24-3
	25	87.6	3.4	547	9	AI604281	AI604281 mt12h04.y
c	26	87.6	3.4	631	9	AI636743	AI636743 ts89e05.x
	27	87.6	3.4	701	17	BH117016	BH117016 RPCI-24-3
	28	87.6	3.4	702	17	AZ768585	AZ768585 1M0568H01
	29	87.4	3.4	470	17	AQ315284	AQ315284 RPCI11-10
	30	87	3.4	671	17	AZ899685	AZ899685 RPCI-24-1
c	31	86.8	3.4	644	17	BH055411	BH055411 RPCI-24-2
	32	86.8	3.4	674	17	BH301997	BH301997 CH230-109
c	33	86.8	3.4	690	10	BB649793	BB649793 BB649793
	34	86.8	3.4	819	17	BH313036	BH313036 CH230-12D
	35	86.6	3.4	617	10	BE467547	BE467547 hz72c07.x
c	36	86.6	3.4	691	17	BH058973	BH058973 RPCI-24-3
	37	86.4	3.4	449	17	AQ003313	AQ003313 RPCI11-20
	38	86.4	3.4	709	17	AZ324218	AZ324218 1M0046M03
	39	86.2	3.4	542	17	AQ455696	AQ455696 HS_5068_B
c	40	86	3.4	788	17	BH045791	BH045791 RPCI-24-3
c	41	85.8	3.3	668	17	AZ420878	AZ420878 1M0198H23
c	42	85.8	3.3	673	10	BE550633	BE550633 7a30g04.x
	43	85.8	3.3	704	17	BH336247	BH336247 CH230-203
	44	85.6	3.3	642	17	AZ802714	AZ802714 2M0061P09
	45	85.6	3.3	712	17	AZ632195	AZ632195 1M0486P09

ALIGNMENTS

RESULT 1
AQ475349
LOCUS AQ475349 705 bp DNA linear GSS 23-APR-1999
DEFINITION CITBI-E1-2609P4.TF CITBI-E1 Homo sapiens genomic clone 2609P4, DNA sequence.
ACCESSION AQ475349
VERSION AQ475349.1 GI:4657468
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 705)
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and Venter,J.C.
TITLE Use of BAC End Sequences from CalTech Libraries for Sequence-Ready Map Building
JOURNAL Unpublished (1997)
COMMENT Other_GSSs: CITBI-E1-2609P4.TR
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.

```
          Seq primer: M13-21
          Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .705
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="2609P4"
                     /clone_lib="CITBI-E1"
                     /sex="male"
                     /cell_type="sperm"
                     /note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI;
                     CalTech Human BAC Library D"
BASE COUNT      235 a    135 c    128 g    207 t
ORIGIN

  Query Match             4.4%;  Score 113.8;  DB 17;  Length 705;
  Best Local Similarity   50.2%;  Pred. No. 5.5e-14;
  Matches  308;  Conservative    0;  Mismatches  302;  Indels    3;  Gaps    1;

Qy    189 TATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAACTAAATAG 248
          | |||   | ||| |      |||||||   ||| | |  || || ||||   ||||
Db     15 TGTAATTCAGAAGAAATCAGGCAAATGGCATATGTTAACTGACTTAAGGGCCGTAAACGC 74

Qy    249 GGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGACTAGCAAA 308
          || | |||   | |   |    |||||   ||  | ||    || ||     |  ||||
Db     75 CGTAATTCAACCCATGGGGCCTCTCCAACCCGGGTTGCCCTCTCCGGCCATGATCCCAAA 134

Qy    309 AAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCTCTAGATGA 368
          |      | |   || | ||| ||   |||   | |||  |||| ||||| |  ||
Db    135 AGATTGGCCTTTAATTATAATTGATCTAAAGGATTGCTTTTTTACCATCCCTCTGGCGGA 194

Qy    369 AGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAGCCAGGAAA 428
          |   | |  |  |   ||||||||||| ||||| |  |||||||  ||| ||||  |
Db    195 GCAGGATTGCGAAAAATTTGCCTTTACTATACCAGCCATAAATAATAAAGAACCAGCCAC 254

Qy    429 ACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATCTTCCAATA 488
          | |     |  || ||  | ||||||||| |         ||| | || |  ||
Db    255 CAGGTTTCAGTGGAAAGTGTTACCTCAGGGAATGCTTAATAGTCCAACTATTTGTCAGAC 314

Qy    489 CACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACCTTAGTCCA 548
          | |  | |  |  ||  ||||  | ||  |    | ||||       |  | ||
Db    315 TTTTGTAGGTCGAGCTCTTCAACCAGTTAGAGACAAGTTTTCAGACTGTTATATTATTCA 374

Qy    549 GTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGACAGGGTAGT 608
          ||||| |||||  || |||   ||||  || | || |||      |  ||||  |
Db    375 TTATATTGATGATATTTTATGTGCTACAGAAACGAGAGATAAATTAATTGACTGTTATAC 434

Qy    609 TTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAGAAATTCCA 668
          || | |  |   ||  |   |||     ||| |  | |   ||   || ||  ||||
Db    435 ATTTCTGCAAGCAGAGGTTGCCAAT---GCAGGACTGGCAATAGCATCTGATAAGATCCA 491

Qy    669 AAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGGAAGTTGCA 728
          ||    | | || |||||  |   | |||    |  |      | ||||   |||   ||
Db    492 AACCTCTACTCCTTTTCATTATTTAGGGATGCAGATAGAAAATAGAAAAATTAAGCCACA 551

Qy    729 AAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAGTTAGTAGG 788
          ||| |||||  |   | ||           ||| | |||||| | || || ||  | ||
Db    552 AAAAATAGAAATAAGAAAAGACACATTAAAAACACTAAATGATTTTCAAAAATTGCTGGG 611

Qy    789 AGTATTAAATTGG 801
          ||   | ||||||
Db    612 AGATATTAATTGG 624


RESULT 2
AQ422201
LOCUS       AQ422201                732 bp    DNA     linear   GSS 23-MAR-1999
DEFINITION  RPCI-11-190013.TV RPCI-11 Homo sapiens genomic clone RPCI-11-190013
            , DNA sequence.
ACCESSION   AQ422201
VERSION     AQ422201.1  GI:4479925
KEYWORDS    GSS.
SOURCE      human.
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 732)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .732
                     /organism="Homo sapiens"
                     /db_xref="GDB:7572924"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-190013"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      246 a    141 c    132 g    212 t      1 others
ORIGIN

  Query Match             4.3%;  Score 110.6;  DB 17;  Length 732;
  Best Local Similarity   49.9%;  Pred. No. 2.7e-13;
  Matches  306;  Conservative    0;  Mismatches  304;  Indels    3;  Gaps    1;

Qy    189 TATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAACTAAATAG 248
          | |||   ||||| |      |||||||  | ||| | |  || || ||||   ||||
Db     15 TGTAATTCAAAAGAAATCATGCAAATGGCGTATGTTAACTGACTTAAGGGCCGTAAACGC 74

Qy    249 GGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGACTAGCAAA 308
          || | |||   | |   |    |||||   ||  | ||    |  ||     |  ||||
Db     75 CGTAATTCAACCCATGGGGCCTCTCCAACCTGGGTTGCCCTCTCTGGCCATGATCCCAAA 134

Qy    309 AAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCTCTAGATGA 368
          |          | |   || | ||| ||   |||   | |||  |||| ||||| |  ||
Db    135 AGACTGGCCTTTAATTATAATTGATCTAAAGGATTGCTTTTTTACCATCCCTCTGGCGGA 194

Qy    369 AGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAGCCAGGAAA 428
          |   | |  |  |   ||||||||||| ||||| |  ||||||||  ||| ||||  |
Db    195 GCAGGATTGCGAAAAATTTGCCTTTACTATACCAGCCATAAATAATAAAGAACCAGCCAC 254

Qy    429 ACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATCTTCCAATA 488
          | |     |  || ||  | |||||||||  |          ||| | || |  ||
Db    255 CAGGTTTCAGTGGAAAGTGTTACCTCAGGGAATGCTTAATAGTCCAACTATTTGTCAGAC 314

Qy    489 CACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACCTTAGTCCA 548
          | |  | |  |  ||  ||||  | ||  |       | ||||       |  | ||
Db    315 TTTTGTAGGTCGAGCTCTTCAACCAGTTAGAGACAAGTTTTCAGACTGTTATATTATTCA 374

Qy    549 GTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGACAGGGTAGT 608
          |||||  |||||  || |||   |||   || |  | |||      |  ||||  |
Db    375 TTATATTGATGATATTTTATGTGCTGCAGAAACAAAAGATAAATTAATTGACTGTTATAC 434

Qy    609 TTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGAAGAGAAATTCCA 668
          || | |  |   ||  |    ||   |  |||  | | |   ||   || ||  ||||
Db    435 ATTTCTGCAAGCAGAGGTTGCCAA---CGCAGGACTGGCAATAGCATCTGATAAGATCCA 491
```

```
Qy     669 AAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAAATGGAAGTTGCA 728
           ||    | | || ||||| |   | |||    |  |       | ||||   |||   ||
Db     492 AACCTCTACTCCTTTTCATTATTTAGGGATGCAGATAGAAAATAGAAAAATTAAGCCACA 551

Qy     729 AAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAGTTAGTAGG 788
           ||| |||||  |   | ||            ||| | ||| || | || || ||  | ||
Db     552 AAAAATAGAAATAAGAAAAGACACATTAAAAACACTAAATTATTTTCAAAAATTGCTGGG 611

Qy     789 AGTATTAAATTGG 801
           ||   | ||||||
Db     612 AGATATTAATTGG 624


RESULT 3
AZ853584
LOCUS       AZ853584                659 bp    DNA     linear   GSS 21-FEB-2001
DEFINITION  2M0156B20R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0156B20 R, DNA sequence.
ACCESSION   AZ853584
VERSION     AZ853584.1  GI:13041853
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 659)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0156  row: B  column: 20
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 659.
FEATURES             Location/Qualifiers
     source          1. .659
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0156B20"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E..coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      221 a    109 c    131 g    198 t
ORIGIN

  Query Match             3.9%;  Score 100.6;  DB 17;  Length 659;
  Best Local Similarity   53.3%;  Pred. No. 3.6e-11;
  Matches  235;  Conservative    0;  Mismatches  204;  Indels    2;  Gaps    1;

Qy     146 AGGAAGCTCCCCCGACCAATCCATACAACACCCCCACATTTGCTATAAAGAAAAAGGATA 205
           | ||| | |||   |||  ||| |  || || || | ||||| ||||||||||||    |
Db       1 ATGAATCGCCCTTCACCTCTCCGTGGAATACTCCTATATTTGTTATAAAGAAAAAAATCA 60

Qy     206 AGAACAAATGGAGAATGCTGATAGATTTTAGGGAACTAAATAGGGTCACTCAGGACTTTA 265
           ||  |||||||||| |  |    ||||| || ||| ||||||  |   |      |    |
Db      61 GGA--AAATGGAGACTTTTACATGATTTGAGAGAAATAAATAAGCAAATGATTTTAATGG 118

Qy     266 CGGAAGTCCAATTAGGAATACCACACCCTGCAGGACTAGCAAAAAGGAAAAGAATTACAG 325
           |||  ||   |||  | |||    || | |   |  | | |        |||| ||
Db     119 GACCAGTACAGAGAGGGCTGCCATTGTCTTCTGCCTTGCCCAGGAATTGGCCCATTATAG 178

Qy     326 TACTGGATATAGGTGATGCATATTTCTCCATACCTCTAGATGAAGAATTTAGGCAGTACA 385
           |  | |||||    |||   | |||||| || ||| || || |  ||  ||
Db     179 TGTTAGATATCAAAGATTGTTTTTTCTCTATTCCTTTATATCATAAAGATACAGTCAGGT 238

Qy     386 CTGCCTTTACTTTACCATCAGTAAATAATGCAGAGCCAGGAAAACGATACATTTATAAGG 445
           |||  |||||| | || ||   |||| |||| || || |  ||  |||     || | ||
Db     239 TTGCTTTTACTGTTCCTTCTCAAAATCATGCTGAACCTGATAAGAGATTTGAATGGACGG 298

Qy     446 TTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATCTTCCAATACACTATGAGACATGTGC 505
           || |||| ||||||  |          ||  | || |  |||     | |      | |
Db     299 TTTTGCCACAGGGAATGGCAAACAGTCCTACTATGTGTCAACTATATGTTGATACAGCCT 358

Qy     506 TAGAACCCTTCAGGAAGGCAAATCCAGATGTGACCTTAGTCCAGTATATGGATGACATCT 565
           |  ||    | || |||  |  |||  | || |  |    ||| || |||||||||||
Db     359 TGAAAGGGGTTAGAAAGAGATTTCCGAAGGTAAAATGTTACCAATACATGGATGACATTC 418

Qy     566 TAATAGCTAGTGACAGGACAG 586
           | || |||     || |  ||
Db     419 TTATTGCTGCCCCCATGGAAG 439


RESULT 4
AZ953622/c
LOCUS       AZ953622                604 bp    DNA     linear   GSS 27-APR-2001
DEFINITION  2M0218O20R Mouse 10kb plasmid UUGC2M library Mus musculus genomic
            clone UUGC2M0218O20 R, DNA sequence.
ACCESSION   AZ953622
VERSION     AZ953622.1  GI:13824849
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 604)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0218  row: O  column: 20
            Seq primer: CACACAGGAAACAGCTATGACC
```

```
          Class: plasmid ends
          High quality sequence stop: 604.
FEATURES             Location/Qualifiers
     source          1. .604
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0218O20"
                     /clone_lib="Mouse 10kb plasmid UUGC2M library"
                     /sex="Female"
                     /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (female) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      190 a    110 c    121 g    183 t
ORIGIN

  Query Match             3.9%;  Score 100.4;  DB 17;  Length 604;
  Best Local Similarity   54.0%;  Pred. No. 3.9e-11;
  Matches  227;  Conservative    0;  Mismatches  191;  Indels    2;  Gaps    1;

Qy     150 AGCTCCCCCGACCAATCCATACAACACCCCCACATTTGCTATAAAGAAAAAGGATAAGAA 209
           ||  ||| | |||    | |  || || || || |||| || ||||||||  |
Db     581 AGAGCCCTCTACCTCATCCTGGAATACACCTACTTTTGTAATTAAGAAAAA--AATCTGG 524

Qy     210 CAAATGGAGAATGCTGATAGATTTTAGGGAACTAAATAGGGTCACTCAGGACTTTACGGA 269
           ||||||| |  |  |    ||| |||| | | | |||     |.        |||  |
Db     523 CAAATGGTGCTTATTACGTGATCTTAGAGCAATTAATGCACAAATGTGCCTGTTTGGGTT 464

Qy     270 AGTCCAATTAGGAATACCACACCCTGCAGGACTAGCAAAAAGGAAAAGAATTACAGTACT 329
           ||| ||   |||  |||||   | | | |  ||| | ||||       ||||| | |  |
Db     463 AGTTCAGGGAGGGCTACCATTGCTTTCTGCCCTACCTAAAAATTGGGAAATTATAATTAT 404

Qy     330 GGATATAGGTGATGCATATTTCTCCATACCTCTAGATGAAGAATTTAGGCAGTACACTGC 389
           || ||    |||   | ||||||| |||||| ||    ||  ||||||      |||
Db     403 AGACATTACAGATTGTTTTTTCTCTATACCTTTATGCCCTCAAGATAGGCAAAGATTTGC 344

Qy     390 CTTTACTTTACCATCAGTAAATAATGCAGAGCCAGGAAAACGATACATTTATAAGGTTCT 449
           |||||  | ||| |  | ||| ||  |||||| |   || |||||   |  ||||| ||
Db     343 ATTTACCATCCCAGCCATTAATCATTTAGAGCCTGATCAAAGATACCAATGGAAGGTCCT 284

Qy     450 GCCTCAGGGATGGAAGGGGTCACCAGCCATCTTCCAATACACTATGAGACATGTGCTAGA 509
           |||||||||  |       ||  |||| |  |||     | ||  |   |  ||  |
Db     283 ACCTCAGGGAATGGCAAATAGCCCCACCATGTATCAACTATATGTGCAATTGGCACTTAA 224

Qy     510 ACCCTTCAGGAAGGCAAATCCAGATGTGACCTTAGTCCAGTATATGGATGACATCTTAAT 569
           |  | || ||      ||||    |    ||  | |  ||||||||||| ||||| ||
Db     223 GTCAGTTAGAAATCATTTTCCATCATTACTTTTGATACTTTATATGGATGATATCTTGAT 164


RESULT 5
BH295203
LOCUS       BH295203                610 bp    DNA     linear   GSS 30-NOV-2001
DEFINITION  CH230-104C16.TJ CHORI-230 Segment 1 Rattus norvegicus genomic clone
            CH230-104C16, DNA sequence.
ACCESSION   BH295203
VERSION     BH295203.1  GI:17207611
KEYWORDS    GSS.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 610)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
            ,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
            Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 EcoRI segment
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: CH230-104C16.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 104  row: C  column: 16
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .610
                     /organism="Rattus norvegicus"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-104C16"
                     /clone_lib="CHORI-230 Segment 1"
                     /sex="Female"
                     /cell_type="Brain"
                     /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
BASE COUNT      204 a    110 c    120 g    176 t
ORIGIN

  Query Match             3.8%;  Score 98.2;  DB 17;  Length 610;
  Best Local Similarity   49.1%;  Pred. No. 1.2e-10;
  Matches  290;  Conservative    0;  Mismatches  298;  Indels    3;  Gaps    1;

Qy     210 CAAATGGAGAATGCTGATAGATTTTAGGGAACTAAATAGGGTCACTCAGGACTTTACGGA 269
           |||||||||  || ||| |||| | || | | | |||| | | | ||||    |
Db      15 CAAATGGAGGGTGTTGACAGATCTCAGAGCAATTAATAAGATTATTCAGCCGATAGGTTC 74

Qy     270 AGTCCAATTAGGAATACCACACCCTGCAGGACTAGCAAAAAGGAAAAGAATTACAGTACT 329
             | ||    || |  |||| ||| |     |  | ||   |      |||| |||  |
Db      75 CCTGCAGCCTGGGAACCCACTGCCTTCTTTGTTGCCTAAGGAGTGGCCTATTATAGTGAT 134

Qy     330 GGATATAGGTGATGCATATTTCTCCATACCTCTAGATGAAGAATTTAGGCAGTACACTGC 389
           ||  ||   | |   | |||| | || ||| |  |||| |    || | |      |||
Db     135 TGACTTAAAAGGTTGTTTTTTCACTATTCCTTTTTATGAGGGCAATAAGGAAAGGTTTGC 194

Qy     390 CTTTACTTTACCATCAGTAAATAATGCAGAGCCAGGAAAACGATACATTTATAAGGTTCT 449
           |||| |  | ||  |  | ||||      | ||   |||| ||||    |  || || ||
Db     195 CTTTTCAGTGCCTACCCTCAATAGACGTCACCCCATAAAAAGATATCAATGGAAAGTCCT 254

Qy     450 GCCTCAGGGATGGAAGGGGTCACCAGCCATCTTCCAATACACTATGAGACATGTGCTAGA 509
           ||| || |||  |          ||  || | |  |||||    | |    ||     | ||
Db     255 GCCACAAGGAATGCTAAATAGCCCTACCTTATGTCAATATTTTGTACAGCAGCCATTGGA 314

Qy     510 ACCCTTCAGGAAGGCAAATCCAGATGTGACCTTAGTCCAGTATATGGATGACATCTTAAT 569
             |  | ||   |  |||  |    |   |    || || ||||||||||| |||
Db     315 GATGATTCGCAAACAATTTCCCCAATCTATTATTTATCATTACATGGATGACATTTTATC 374

Qy     570 AGCTAGTGACAGGACAGACCTGGAACATGACAGGGTAGTTTTACAGTCAAAGGAACTCTT 629
```

```
Db    375 GTCTGATTCAGATATAAATAT---CCTGGAAAGAATGTTTGATGAAGTGAAAAAAACCTT 431

Qy    630 GAATAGCATAGGGTTTTCTACCCCAGAAGAGAAATTCCAAAAAGATCCCCCATTTCAATG 689

Db    432 ACCTTGCTGGGGATTACGAATCGCCCCTGAGAAAATACAAAGAGGGGATTCTATTAATTA 491

Qy    690 GATGGGGTACGAATTGTGGCCAACAAAATGGAAGTTGCAAAAGATAGAGTTGCCACAAAG 749

Db    492 TCTAGGGTATAAAATAAGTCTACAAAGGATTAGGCCACAAAAGGTGCAAATCAAAAGAAA 551

Qy    750 AGAGACCTGGACAGTGAATGATATACAGAAGTTAGTAGGAGTATTAAATTG 800

Db    552 TCAATTAAGGACACTTAATGATTTTCAGAAACTACTGCGAGATATTAACTG 602


RESULT 6
AZ362320
LOCUS       AZ362320               706 bp    DNA     linear   GSS 02-OCT-2000
DEFINITION  1M0107E20F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0107E20 F, DNA sequence.
ACCESSION   AZ362320
VERSION     AZ362320.1  GI:10476020
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 706)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0107  row: E  column: 20
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 706.
FEATURES             Location/Qualifiers
     source          1. .706
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0107E20"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      230 a    138 c    128 g    210 t
ORIGIN

  Query Match             3.8%;  Score 97.2;  DB 17;  Length 706;
  Best Local Similarity   47.5%;  Pred. No. 2e-10;
  Matches  322;  Conservative    0;  Mismatches  353;  Indels    3;  Gaps    1;

Qy    157 CCGACCAATCCATACAACACCCCCACATTTGCTATAAAGAAAAAGGATAAGAACAAATGG 216

Db     11 CCTTCTAACTCACCCTGGAATACACCTATTTTCGTAATTAAGAAAAAATCTGGCAAATGG 70

Qy    217 AGAATGCTGATAGATTTTAGGGAACTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAA 276

Db     71 CGCTTATTACATGATCTTAGAGCAATTAATGCGCAAATGCATCTGTTTGGACCAGTTCAG 130

Qy    277 TTAGGAATACCACACCCTGCAGGACTAGCAAAAAGGAAAAGAATTACAGTACTGGATATA 336

Db    131 CAAGGGCTACCATTACTTTCTGCCTTACCCAAAAATTGGGAAATTATAATCCTAGACATT 190

Qy    337 GGTGATGCATATTTCTCCATACCTCTAGATGAAGAATTTAGGCAGTACACTGCCTTTACT 396

Db    191 AAAAATTGCTTTTTCTCTATACCTTTATGCCCTCAAGATAGGCAAAGATTTGCATTTACG 250

Qy    397 TTACCATCAGTAAATAATGCAGAGCCAGGAAAACGATACATTTATAAGGTTCTGCCTCAG 456

Db    251 ATCCCAGCCATTAATTACTTAGAGCCAGATCAAAAATACCAATGGAAGGTCCTACCTCAG 310

Qy    457 GGATGGAAGGGGTCACCAGCCATCTTCCAATACACTATGAGACATGTGCTAGAACCCTTC 516

Db    311 GGGATGGCGAATAGCCCCACCATGTGTCAACTGTATGTACAATTGGCACTTAAGTCAGTT 370

Qy    517 AGGAAGGCAAATCCAGATGTGACCTTAGTCCAGTATATGGATGACATCTTAATAGCTAGT 576

Db    371 AGAGAACATTTTCCATCATTACAGGTGATAATTTATATGGATGACATTTTGATTTGCCAT 430

Qy    577 GACAGGACAGACCTGGAACATGACAGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGC 636

Db    431 AAAAATTCAGAGTTATTGCAAGATGCATACCTTATATTAATAAAAACGTTAGGGCAATGG 490

Qy    637 ATAGGGTTTTCTACCCCAGAAGAGAAATTCCAAAAAGATCCCCCATTTCAATGGATGGGG 696

Db    491 GGATTGCAGGTGGCCACCGAA---AAGGTGCAAGTTGCTCGAATGGGAGCCTTCCTAGGG 547

Qy    697 TACGAATTGTGGCCAACAAAATGGAAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACC 756

Db    548 TCACTTATTTATCCTGACAAAATTGTTCCTCAAAAATTAGAGATTCGCAAAGATCAACTA 607

Qy    757 TGGACAGTGAATGATATACAGAAGTTAGTAGGAGTATTAAATTGGGCAGCTCAAATTTAT 816

Db    608 CATACCTTGAATGATTTTCAAAAATTGTTAGGAGATATTAATTGGGCTGAGACCCTTTTT 667

Qy    817 CCAGGTATAAAAACCAAA 834

Db    668 GAAAATTCCATCAGCAGA 685


RESULT 7
AQ008834
LOCUS       AQ008834               548 bp    DNA     linear   GSS 14-APR-1999
DEFINITION  RPCI11-20P12.TPD RPCI-11 Homo sapiens genomic clone RPCI-11-20P12,
            DNA sequence.
ACCESSION   AQ008834
VERSION     AQ008834.1  GI:3129605
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 548)
```

```
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences for Sequence-Ready Map Building (1998)
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: RPCI11-20P12.TV
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .548
                     /organism="Homo sapiens"
                     /db_xref="GDB:7507667"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-20P12"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      165 a    112 c     96 g    175 t
ORIGIN

  Query Match             3.7%;  Score 95.4;  DB 17;  Length 548;
  Best Local Similarity   53.3%;  Pred. No. 4.5e-10;
  Matches  201;  Conservative    0;  Mismatches  176;  Indels    0;  Gaps    0;

Qy    197 AAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAACTAAATAGGGTCACTC 256
          ||||| |      ||||||| | ||  | |  ||||| ||||   | |||   || | |
Db      7 AAAAGAAATCATGCAAATGGCGGATAGTAACCGATTTAAGGGCCATTAATGCTGTAATTA 66

Qy    257 AGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGACTAGCAAAAAGGAAAA 316
          |     |   ||  ||||||   || || ||   ||||||   | || | |||
Db     67 AACCTATGGGGGCCGTCCAACCCGGCATGCCTGCCCCTGCTTTAATACCTAAAGATTGGC 126

Qy    317 GAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCTCTAGATGAAGAATTTA 376
            | |||  | ||| |    ||     | |||    ||  || || || ||     |
Db    127 CTCTCATAGTTATTGATCTTAAAGAGTTTTTTTTTCATATCGCTTTACATAAATCGGATT 186

Qy    377 GGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAGCCAGGAAAACGATACA 436
          |  |  |   ||| |||||| |||||||  | ||||||   ||||| | |   || ||
Db    187 GTGAAAAATTTGCTTTTACTGTACCATCTATCAATAATCAGGAGCCTGTAGCTCGTTATC 246

Qy    437 TTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATCTTCCAATACACTATGA 496
          |  || || || |||||||||  |         ||  | |||| |||      | |
Db    247 AATGGAAAGTACTTCCTCAGGGAATGCTAAATAGCCCTACAATCTGCCAGCTTTATGTTG 306

Qy    497 GACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACCTTAGTCCAGTATATGG 556
          ||||  |||||   |||  |  |       |  ||| || |      |  | || |||||
Db    307 GACAAGTGCTTTCACCAGTTTGAGCCCGATTTCCCGAGGCCTATATTCTTCATTATATTG 366

Qy    557 ATGACATCTTAATAGCT 573
          |||| || |||||  ||
Db    367 ATGATATTTTAATTTCT 383


RESULT 8
AZ720063/c
LOCUS       AZ720063                641 bp    DNA     linear   GSS 24-JAN-2001
DEFINITION  RPCI-24-155P5.TJ RPCI-24 Mus musculus genomic clone RPCI-24-155P5,
            DNA sequence.

ACCESSION   AZ720063
VERSION     AZ720063.1  GI:12461383
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 641)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-155P5.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 155  row: P  column: 5
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .641
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-155P5"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT      194 a    123 c    117 g    207 t
ORIGIN

  Query Match             3.7%;  Score 95.2;  DB 17;  Length 641;
  Best Local Similarity   52.8%;  Pred. No. 5.1e-10;
  Matches  229;  Conservative    0;  Mismatches  203;  Indels    2;  Gaps    1;

Qy    166 CCATACAACACCCCCACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTG 225
          || |  || || |||| |||||  || ||||||||  || ||   ||  ||||  | |||
Db    570 CCTTGGAATACACCCATATTTGTGATTAAGAAAAA--ATCAGGCAAATGGGAGGCTCCTG 513

Qy    226 ATAGATTTTAGGGAACTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATA 285
            ||| || | |   | ||||     |  |||    |        |  |    ||  |
Db    512 CATGATCTTCGAGCCATTAATAATCAAATGCAGATTATGGGCCCCATATAGAGGGGTCTG 453

Qy    286 CCACACCCTGCAGGACTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCA 345
          || |  | | |  |  |  | |||  |      ||||  | |  | |||||    |||
Db    452 CCCCTACTTTCCAGTTTGCCCAAAGAGTGGCCTATTATAATCATTGATATTAAGGATTGT 393

Qy    346 TATTTCTCCATACCTCTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCA 405
          | |||||| || || || |          |||  || ||  ||||||||||  | || ||
Db    392 TTTTTCTCTATTCCCCTTGCATTGTGTGATAGTGAGCACTTTGCCTTTACTCTTCCTTCT 333

Qy    406 GTAAATAATGCAGAGCCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAG 465
          |||||||||| ||||||  |  ||| | ||    |  |  ||| | ||||| ||   |
Db    332 GTAAATAATGAAGAGCCTGATAAATGGTATCAATGGATAGTTTTACCTCAAGGGATGGCA 273

Qy    466 GGGTCACCAGCCATCTTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCA 525
                ||  | || |  |||     | |  || | |  | | ||||  || |  | |||
Db    272 AACAGTCCTACTATGTGTCAACTTTTTGTAGGAGAAGCCCCACAACCTGTCCGTGATGCA 213
```

```
Qy     526 AATCCAGATGTGACCTTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACA 585
           ||| |  | |   |||| || ||| | |||||  || |  | ||     | |  | |
Db     212 TTCCCAAAATTAAGGATAGTACATTATGTAGATGATGTCCTTCTTGCCTCAAAGAATAAA 153

Qy     586 GACCTGGAACATGA 599
           ||      | ||||
Db     152 GAAAGTTTAGATGA 139


RESULT 9
AZ589335/c
LOCUS       AZ589335               704 bp    DNA     linear   GSS 13-DEC-2000
DEFINITION  1M0398D17F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0398D17 F, DNA sequence.
ACCESSION   AZ589335
VERSION     AZ589335.1  GI:11711525
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 704)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0398  row: D  column: 17
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 704.
FEATURES             Location/Qualifiers
     source          1. .704
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0398D17"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      211 a    130 c    131 g    232 t
ORIGIN
```

```
  Query Match             3.7%;  Score 94.8;  DB 17;  Length 704;
  Best Local Similarity   48.9%;  Pred. No. 6.4e-10;
  Matches  312;  Conservative    0;  Mismatches  322;  Indels    4;  Gaps    2;

Qy     177 CCCCACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAG 236
           | ||   |||  ||   |  ||||  ||  |   |||||| |  |  |    ||| ||||
Db     697 CACCTATTTTCGTAATTAAGAAAAAAATCTGGCAAAATGGCGCTTATTACATGATCTTAG 638

Qy     237 GGAACTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGC 296
           | | | |||  |   |  ||    |||     |||  ||   |||  |||||   | | |
Db     637 AGCAATTAATGCGCAAATGCATCTGTTTGGACCAGTTCAGCAAGGGCTACCATTACTTTC 578

Qy     297 AGGACTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCAT 356
           |   || | ||||       |||||  | | || || ||     ||   | ||||| ||
Db     577 TGCCTTACCCAAAAATTGGGAAATTATAATCCTAGACATTAAAAATTGCTTTTTCTCTAT 518

Qy     357 ACCTCTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGC 416
           |||| ||       ||  ||||||      ||| |||||  | ||| |  | ||| |
Db     517 ACCTTTATGCCCTCAAGATAGGCAAAGATTTGCATTTACGATCCCAGCCATTAATTACTT 458

Qy     417 AGAGCCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGC 476
           ||||||||   || |||||   |  ||||| || ||||||||   |  |      ||  |
Db     457 AGAGCCAGATCAAAGATACCAATGGAAGGTCCTACCTCAGGGGATGGCGAATAGCCCCAC 398

Qy     477 CATCTTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGT 536
           ||| |  |||     | |   |   | ||  |  |  | ||  |      ||||    |
Db     397 CATGTGTCAACTGTATGTACAATTGGCACTTAAGTCAGTTAGAGAACATTTTCCATCATT 338

Qy     537 GACCTTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACA 596
           |  |    |||||||||||||| || ||      | | |   ||||  |    ||
Db     337 ACAGGTGATAATTTATATGGATGACATTTTGATTTGCCATAAAAATTCAGAGTTATTGCA 278

Qy     597 TGACAGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATAGGGTTTTCTACCCCAGA 656
           | | |      ||| | | | |  |||  | |  |     |  |        |||
Db     277 --AGATGCATACCTTATATTAATAAAAACGTTAGGGCAATGGGGATTGCAGGTGGCCACC 220

Qy     657 AGAGAAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCAACAAA 716
           | |  | | |                |   |||  ||  |||    | | |||||
Db     219 GAAAAGGTGCAAGTTGCTCGAATGGGAGACTTCCTAGGGTCACTTATTTATCCTGACAAA 160

Qy     717 ATGGAAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACA 776
           || |    |  |||  | ||| |||| |  |  |   || |  ||  |||||||| | ||
Db     159 ATTGTTCCTCAAAATTAGAGATTTGCAAAGATCAACTACAT--ACCTTGAATGATTTTCA 102

Qy     777 GAAGTTAGTAGGAGTATTAAATTGGGCAGCTCAAATTT 814
           | || ||  ||||||  | ||||||    | |   |||
Db     101 AAAATTGTTAGGAGATATTAATTGGCTGACACCCTTTT 64


RESULT 10
AI633818/c
LOCUS       AI633818               586 bp    mRNA    linear   EST 14-DEC-1999
DEFINITION  tt28h04.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2242135 3'
            similar to gb:M14123_cds4 RETROVIRUS-RELATED POL POLYPROTEIN (HUMAN
            );, mRNA sequence.
ACCESSION   AI633818
VERSION     AI633818.1  GI:4685148
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 586)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
```

```
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 611   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 199.
FEATURES              Location/Qualifiers
     source           1. .586
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:2242135"
                      /clone_lib="NCI_CGAP_GC6"
                      /tissue_type="pooled germ cell tumors"
                      /lab_host="DH10B"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA
                      from the normalized library NCI_CGAP_GC4 was prepared, and
                      ss circles were made in vitro. Following HAP purification,
                      this DNA was used as tracer in a subtractive hybridization
                      reaction. The driver was PCR-amplified cDNAs from a pool
                      of 5,000 clones made from the same library (cloneIDs
                      1257096-1258631, 1469064-1470983, and 1475592-1476743).
                      Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      172 a    118 c    117 g    179 t
ORIGIN

  Query Match             3.6%;  Score 93.4;  DB 9;  Length 586;
  Best Local Similarity   51.3%;  Pred. No. 1.2e-09;
  Matches  217;  Conservative    0;  Mismatches  206;  Indels    0;  Gaps    0;

Qy     180 CACATTTGCTATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGA 239
           | | | || |   |   |||||||      |||||||   | | |  | || || ||||
Db     545 CTCTTGTGTTGGTATTCAAAAGGAATCAGGCAAATGGCATACGTTACTGGACTTAAGGGC 486

Qy     240 ACTAAATAGGGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGG 299
           ||||   || | |||   | |   |   |||||   ||  | ||    || ||
Db     485 TGTAAACGCCGTAATTCAACCCATGGGGCCTCTCCAACCCGGGTTGCCCTCTCCGGCCAT 426

Qy     300 ACTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACC 359
           |  |||||          | | | || | ||| || | |||   | |||  |||| ||
Db     425 GATCCCAAAAGATTGGCCTTTAATAATAATTGATCTAAGGGATTGCTTTTTACCCATCCC 366

Qy     360 TCTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGA 419
           ||| |  ||  |   | |  |  |  ||||||||||| ||||| |  ||||||||  |||
Db     365 TCTGGCAGAGCAGGATTGTGAAAAATTTGCCTTTACTATACCAGCCATAAATAATAAAGA 306

Qy     420 GCCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCAT 479
           ||||  |   | |     |  || ||  | |||||||||  |         ||| | ||
Db     305 ACCAGCCACCAGGTTTCAGTGGAAAGTGTTACCTCAGGGAATGCTTAATAGTCCAACTAT 246

Qy     480 CTTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGAC 539
           |  ||      | |  | |  |  ||  ||||  | ||  |      | ||||
Db     245 TTGTCAGACTTTTGTAGGTCGAGCTCTTCAACCAGTGAGAGAAAAGTTTTCAGACTGTTA 186

Qy     540 CTTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGA 599
           |  | || ||||| |||| || |||   |||   || | || |||      |  |||
Db     185 TATTATTCATTATATTGATGATATTTTATGTGCTGCAGAAACGAAAGATAAATTAATTGA 126

Qy     600 CAG 602
           | |
Db     125 CTG 123


RESULT 11
AZ430794
LOCUS       AZ430794                 620 bp    DNA     linear   GSS 03-OCT-2000
DEFINITION  1M0215K23F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0215K23 F, DNA sequence.
ACCESSION   AZ430794
VERSION     AZ430794.1  GI:10554807
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 620)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0215  row: K  column: 23
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 620.
FEATURES              Location/Qualifiers
     source           1. .620
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="UUGC1M0215K23"
                      /clone_lib="Mouse 10kb plasmid UUGC1M library"
                      /sex="Male"
                      /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                      /note="Vector: PWD42nv; Purified genomic DNA from M.
                      musculus C57BL/6J (male) was obtained  from the Jackson
                      Laboratory Mouse DNA Resource
                      (http://www.jax.org/resources/documents/dnares/). The DNA
                      was hydrodynamically sheared by repeated passage through a
                      0.005 inch orifice at constant velocity. The sheared DNA
                      was blunt end-repaired with T4 DNA polymerase and T4
                      polynucleotide kinase. Adaptor oligonucleotides were
                      ligated to the blunt ends in high  molar excess. The
                      adaptored DNA was purified and size-selected for a 9.5 to
                      10.5 kb range using preparative agarose gel
                      electrophoresis. Vector DNA was prepared from a derivative
                      of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                      inducible derivative of plasmid R1. The vector was ligated
                      with adaptors complementary to the insert adaptors and
                      purified. The sheared, adaptored mouse DNA was annealed to
                      adaptored vector DNA, and transformed into
                      chemically-competent E. coli XL10-Gold (Stratagene) cells
                      and selected for ampicillin resistance."
BASE COUNT      195 a    112 c    109 g    204 t
ORIGIN

  Query Match             3.6%;  Score 93.2;  DB 17;  Length 620;
  Best Local Similarity   50.0%;  Pred. No. 1.4e-09;
  Matches  261;  Conservative    0;  Mismatches  258;  Indels    3;  Gaps    1;

Qy     280 GGAATACCACACCCTGCAGGACTAGCAAAAAGGAAAAGAATTACAGTACTGGATATAGGT 339
           ||  | || |  | | |  |  |  | |||  |      |||| | |  | |||||
Db       3 GGTCTGCCTCTACTTTCTAGTTTGCCCAAAGAGTGGCCTATTATAATCATTGATATTAAG 62

Qy     340 GATGCATATTTCTCCATACCTCTAGATGAAGAATTTAGGCAGTACACTGCCTTTACTTTA 399
           |||   | |||||| || || || |          |||  ||  |  ||||||||||  |
Db      63 GATTGTTTTTTCTCTATTCCCCTTGCACTGTGTGATAGCAAGCGCTTTGCCTTTACTCTT 122

Qy     400 CCATCAGTAAATAATGCAGAGCCAGGAAAACGATACATTTATAAGGTTCTGCCTCAGGGA 459
           || || ||||||||||  ||||||  |  |||||  |      |  |  |||  | |||||  ||
Db     123 CCTTCTGTAAATAATGAAGAGCCTGATAAACGGTGTCAGTGGATAGTTTTACCTCAAGGG 182
```

```
Qy     460 TGGAAGGGGTCACCAGCCATCTTCCAATACACTATGAGACATGTGCTAGAACCCTTCAGG 519
Db     183 ATGGCAAACAGTCCTAATATGTGTCAACTTTTTGTAGGAGGAGCCCTACAACCTGTCCAT 242

Qy     520 AAGGCAAATCCAGATGTGACCTTAGTCCAGTATATGGATGACATCTTAATAGCTAGTGAC 579
Db     243 AATGCATTCCCAAAATTAAGGATAATACATTATATAGATGATGTCCTTCTTGCCTCAAAG 302

Qy     580 AGGACAGACCTGGAACATGACAGGGTAGTTTTACAGTCAAAGGAACTCTTGAATAGCATA 639
Db     303 AATAAAGAAAGTTTAGATGA---AGCCTATATTAAGCTAGTAAAATAATTAGAAATGAAA 359

Qy     640 GGGTTTTCTACCCCAGAAGAGAAATTCCAAAAAGATCCCCCATTTCAATGGATGGGGTAC 699
Db     360 CAATTATTTATAGCTCCTGATAAAGTTCAAATGGGAAATTTAGGAGAGTTCTTGGGTGCT 419

Qy     700 GAATTGTGGCCAACAAAATGGAAGTTGCAAAAGATAGAGTTGCCACAAAGAGAGACCTGG 759
Db     420 CGAATTACTCCACATTTTATTACACCTCAAAAGATAGAATTACATAAAGATCATTTGAAA 479

Qy     760 ACAGTGAATGATATACAGAAGTTAGTAGGAGTATTAAATTGG 801
Db     480 ACATTGAATGATTTCCAGAAGTTGTTGGGAGATATAAATTAG 521


RESULT 12
AQ582156
LOCUS       AQ582156                 629 bp    DNA     linear   GSS 07-JUN-1999
DEFINITION  RPCI-11-450N24.TV RPCI-11 Homo sapiens genomic clone RPCI-11-450N24
            , DNA sequence.
ACCESSION   AQ582156
VERSION     AQ582156.1  GI:5009266
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 629)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .629
                     /organism="Homo sapiens"
                     /db_xref="GDB:7672751"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-450N24"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      212 a    112 c    114 g    190 t      1 others
ORIGIN

  Query Match             3.6%;  Score 93.2;  DB 17;  Length 629;
  Best Local Similarity   49.8%;  Pred. No. 1.4e-09;
  Matches  264;  Conservative    0;  Mismatches  263;  Indels    3;  Gaps    1;

Qy     272 TCCAATTAGGAATACCACACCCTGCAGGACTAGCAAAAAGGAAAAGAATTACAGTACTGG 331
Db      94 TCCAACCTGGGTTGCCCTCTCCAGCCATGATCCCAAAAGATTGGCCTTTAATTATAATTG 153

Qy     332 ATATAGGTGATGCATATTTCTCCATACCTCTAGATGAAGAATTTAGGCAGTACACTGCCT 391
Db     154 ATCTAAAGGATTGCTTTTTTACCATCCCTCTGGCAGAGCAGGATTGCGAAAAATTCGCCT 213

Qy     392 TTACTTTACCATCAGTAAATAATGCAGAGCCAGGAAAACGATACATTTATAAGGTTCTGC 451
Db     214 TTACTATACCAGCCATAAATAATAAAGAACCAGCCACCAGGTTTCAGTGGAAAGTGTTAC 273

Qy     452 CTCAGGGATGGAAGGGGTCACCAGCCATCTTCCAATACACTATGAGACATGTGCTAGAAC 511
Db     274 CTCAGGGAATGCTTAATAGTCCAACTATTTGTCAGACTTTTGTAGGTCGAGCTCTTCAAC 333

Qy     512 CCTTCAGGAAGGCAAATCCAGATGTGACCTTAGTCCAGTATATGGATGACATCTTAATAG 571
Db     334 CAGTTAGAGAAAAGTTTTCAGACTGTTATATTATTCATTATATTGATGATATTTTATGTG 393

Qy     572 CTAGTGACAGGACAGACCTGGAACATGACAGGGTAGTTTTACAGTCAAAGGAACTCTTGA 631
Db     394 CTGCAGAAACGAGAGATAAATTAATTGACTGTTATACATTTCTGCAAGCAGAGGTTGCCA 453

Qy     632 ATAGCATAGGGTTTTCTACCCCAGAAGAGAAATTCCAAAAAGATCCCCCATTTCAATGGA 691
Db     454 AT---GCTGGACTGGCAATAGCATCTGATAAGATCCAAACCTCTACTCCTTTTCATTATT 510

Qy     692 TGGGGTACGAATTGTGGCCAACAAAATGGAAGTTGCAAAAGATAGAGTTGCCACAAAGAG 751
Db     511 TAGGGATGCAGATAGAAAATAGAAAAATTAAGCCACAAAAAATAGAAATAAGAAAAGACA 570

Qy     752 AGACCTGGACAGTGAATGATATACAGAAGTTAGTAGGAGTATTAAATTGG 801
Db     571 CATTAAAAACACTAAATGATTTTTAAAAATTACTAGGAGATATTAATTGG 620


RESULT 13
AQ401360
LOCUS       AQ401360                 480 bp    DNA     linear   GSS 13-MAR-1999
DEFINITION  HS_5064_A2_C04_SP6E RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=640 Col=8 Row=E, DNA sequence.
ACCESSION   AQ401360
VERSION     AQ401360.1  GI:4412703
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 480)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
```

```
           Plate: 640  row: E  column: 8
           Seq primer: SP6
           Class: BAC ends
           High quality sequence stop: 480.
FEATURES             Location/Qualifiers
     source          1. .480
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=640 Col=8 Row=E"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT      144 a     94 c     95 g    147 t
ORIGIN

  Query Match             3.6%;  Score 93;  DB 17;  Length 480;
  Best Local Similarity   50.6%;  Pred. No. 1.4e-09;
  Matches  225;  Conservative    0;  Mismatches  220;  Indels    0;  Gaps    0;

Qy    189 TATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAACTAAATAG 248
          | |||   ||||| |      |||||| | ||| | |  || || |||   ||||
Db     11 TGTAATTCAAAAGAAATCAGGCAAATGGCGTATGTTAACTGACTTAAGGGCTGTAAACGC 70

Qy    249 GGTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGACTAGCAAA 308
           || | |||   | |   |    |||||  ||  | ||   || ||     |  ||||
Db     71 CGTAATTCAACCCATGGGGCCTCTCCAACCCGGGTTGCCCTCTCCGGCCATGATCCCAAA 130

Qy    309 AAGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCTCTAGATGA 368
          |          | |   || | ||| ||   |||   | |||  |||| ||||| |  ||
Db    131 AGATTGGCCTTTAATTATAATTGATCTAAAGGATTGCTTTTTTACCATCCCTCTGGCAGA 190

Qy    369 AGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAGCCAGGAAA 428
            |  | |  |  |  ||||||||| ||||| |  |||||||  ||| ||||  |
Db    191 GCAGGATTGCGAAAAATTTGCCTTTACTATACCAGCCATAAATAATAAAGAACCAGCCAC 250

Qy    429 ACGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATCTTCCAATA 488
            | |     |  || ||  | |||||||||  |         ||| | || |  ||
Db    251 CAGGTTTCAGTGGAAAGTGTTACCTCAGGGAATGCTTAATAGTCCAACTATTTGTCAGAC 310

Qy    489 CACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACCTTAGTCCA 548
            | |  | |  |  ||  ||||  | ||  |      | ||||        |  | ||
Db    311 TTTTGTAGGTCGAGCTCTTCAACCAGTTAGAGAAAAGTTTTCAGACTGTTATATTATTCA 370

Qy    549 GTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGACAGGGTAGT 608
          | ||| ||||| || |||   |||   || | || |||      |  |||| |
Db    371 TTGTATTGATGATATTTTATGTGCTGCAGAAACGAAAGATAAAATAATTGACTGTTATAC 430

Qy    609 TTTACAGTCAAAGGAACTCTTGAAT 633
            || |   ||   ||| |  | |||
Db    431 ATTTCTTGCAGCAGAAGTTGTCAAT 455


RESULT 14
AZ345218/c
LOCUS       AZ345218                 708 bp    DNA     linear   GSS 29-SEP-2000
DEFINITION  1M0079H11R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0079H11 R, DNA sequence.
ACCESSION   AZ345218
VERSION     AZ345218.1  GI:10424455
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 708)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
```

```
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0079  row: H  column: 11
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 708.
FEATURES             Location/Qualifiers
     source          1. .708
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0079H11"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      211 a    138 c    128 g    231 t
ORIGIN

  Query Match             3.6%;  Score 92.4;  DB 17;  Length 708;
  Best Local Similarity   52.0%;  Pred. No. 2.1e-09;
  Matches  260;  Conservative    0;  Mismatches  231;  Indels    9;  Gaps    2;

Qy     70 CAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAAAAGATGGAA 129
          |||||||| ||| |    || ||| ||   |||   | |||| |     ||      |||
Db    670 CAGTGGCCTTTACCCTCTGAGAAGTTAAAAGCAGCTAAAGAATTAGTACAA------GAA 617

Qy    130 AAGGATGGTCAGTTGGAGGAAGCTCCCCCGACCAATCCATACAACACCCCCACATTTGCT 189
           |  | |  | |  | | | ||  ||| ||      || |  || ||||| |  |||| |
Db    616 CAACAAGACCTGGGGCATGTAGACCCCACGCAATCACCCTGGAATACCCCTATTTTTGTT 557

Qy    190 ATAAAGAAAAAGGATAAGAACAAATGGAGAATGCTGATAGATTTTAGGGAACTAAATAGG 249
          |||||||  |||         |||||||| || |    ||  | || |   | |||
Db    556 GTAAAGAAGAAG---TCAGGTAAATGGAGATTGTTACATGACCTAAGAGCCATTAATGCA 500

Qy    250 GTCACTCAGGACTTTACGGAAGTCCAATTAGGAATACCACACCCTGCAGGACTAGCAAAA 309
            |  || | | |      || |||  |||| | || |  |   |||  ||| | ||
Db    499 CAAATGCAAGTCATGTGTCCTGTACAAAGAGGACTCCCCCTACTCTCAGCTCTACCTAAG 440

Qy    310 AGGAAAAGAATTACAGTACTGGATATAGGTGATGCATATTTCTCCATACCTCTAGATGAA 369
              || |||| |||  |||||||    |||   |  || ||||| ||  || || |
Db    439 GATTGGAGGATTATAGTGGTGGATATTAAGGATTGTTTCTTTTCCATTCCATTAAATAAG 380

Qy    370 GAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAGCCAGGAAAA 429
           ||  ||  |       ||| ||||| || || ||  | ||| ||   ||||| |  |||
```

```
Db    379 AAAGATAAACCTAGATTTGCATTTACCTTGCCTTCTATTAATCATATGGAGCCTGATAAA 320

Qy    430 CGATACATTTATAAGGTTCTGCCTCAGGGATGGAAGGGGTCACCAGCCATCTTCCAATAC 489

Db    319 AGGTATCAATGGCATGTATTGCCCCAAGGAATGGCAAATAGCCCTACCATATGTCATTTA 260

Qy    490 ACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAATCCAGATGTGACCTTAGTCCAG 549

Db    259 TATGTAGGAAAAGCTTTGCAACCAGTTAGAGATGGTTTCCCCGCCTTAAAAAATTGTCAT 200

Qy    550 TATATGGATGACATCTTAAT 569
          ||||||||||| ||  ||||
Db    199 TATATGGATGATATTGTAAT 180


RESULT 15
BH287730
LOCUS       BH287730                511 bp    DNA     linear   GSS 30-NOV-2001
DEFINITION  CH230-181B17.TV CHORI-230 Segment 1 Rattus norvegicus genomic clone
            CH230-181B17, DNA sequence.
ACCESSION   BH287730
VERSION     BH287730.1  GI:17200138
KEYWORDS    GSS.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 511)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
            ,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
            Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 EcoRI segment
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 181  row: B  column: 17
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .511
                     /organism="Rattus norvegicus"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-181B17"
                     /clone_lib="CHORI-230 Segment 1"
                     /sex="Female"
                     /cell_type="Brain"
                     /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
BASE COUNT      157 a     89 c    107 g    158 t
ORIGIN

  Query Match             3.6%;  Score 92;  DB 17;  Length 511;
  Best Local Similarity   53.3%;  Pred. No. 2.4e-09;
  Matches  194;  Conservative    0;  Mismatches  170;  Indels    0;  Gaps    0;

Qy    210 CAAATGGAGAATGCTGATAGATTTTAGGGAACTAAATAGGGTCACTCAGGACTTTACGGA 269
          |||||||||  || ||| |||| | || | | | |||| | | | ||||    |
Db      3 CAAATGGAGGGTGTTGACAGATCTCAGAGCAATTAATAAGATTATTCAGCTGATGGGTTC 62

Qy    270 AGTCCAATTAGGAATACCACACCCTGCAGGACTAGCAAAAAGGAAAAGAATTACAGTACT 329
            | ||     || || ||||  ||| |      |  | ||           |||| |||  |
Db     63 CCTGCAGCCTGGGATCCCACTGCCTTCTTTGTTGCCTAAGGAATGGCCTATTATAGTGAT 122

Qy    330 GGATATAGGTGATGCATATTTCTCCATACCTCTAGATGAAGAATTTAGGCAGTACACTGC 389
            ||  ||   |||   | |||| | || |||||| |||||||   || | |     |||
Db    123 TGACTTAAAAGATTGTTTTTTCACTATTCCTCTACATGAAGACGATAAGGAAAGGTTTGC 182

Qy    390 CTTTACTTTACCATCAGTAAATAATGCAGAGCCAGGAAAACGATACATTTATAAGGTTCT 449
          ||||  |  | ||  ||  | ||||  |   ||||   |||| ||||    |  || ||  ||
Db    183 CTTTTCAGTGCCTTCCCTCAATAGAGGTCAGCCCATAAAAAGATATCAATGGAAAGTCCT 242

Qy    450 GCCTCAGGGATGGAAGGGGTCACCAGCCATCTTCCAATACACTATGAGACATGTGCTAGA 509
          ||| ||||||  |  |       ||  || | |  |||||   | |   |||      | ||
Db    243 GCCGCAGGGAATGTTGACTAGCCCTACCTTGTGTCAATATTTTGTACAACAGTCATTGGA 302

Qy    510 ACCCTTCAGGAAGGCAAATCCAGATGTGACCTTAGTCCAGTATATGGATGACATCTTAAT 569
              |  | ||   |  |||  |     |   |    || || |||||||| || ||| |
Db    303 GAAGATTCGCAAACAATTTCCCCAGTCTATTATTTATCATTACATGGATGATATTTTATT 362

Qy    570 AGCT 573
            ||
Db    363 GTCT 366


Search completed: November 16, 2002, 13:07:06
Job time : 2396.97 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 16, 2002, 07:31:18 ; Search time 4512.85 Seconds
(without alignments)
17031.470 Million cell updates/sec

Title: M33262_COPY_6695_9335
Perfect score: 2641
Sequence: 1 TCTAGAAGCATGCTATAACA.....ACTGACCTACCTACAATATG 2641

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2641	100.0	10535	14	SIVMM239	M33262 Simian (mac
2	2618	99.1	2799	14	SIVMM316ZZ	L22823 Simian immu
3	2614.8	99.0	2799	14	SIVMM316ZW	L22820 Simian immu
4	2613.2	98.9	2799	14	SIVMM316ZS	L22816 Simian immu
5	2613.2	98.9	2799	14	SIVMM316ZU	L22818 Simian immu
6	2612.2	98.9	10279	14	AY033233	AY033233 Simian im
7	2611.6	98.9	2799	14	SIVMM316ZY	L22822 Simian immu
8	2610	98.8	2799	14	SIVMM316ZT	L22817 Simian immu
9	2610	98.8	2799	14	SIVMM316ZV	L22819 Simian immu
10	2610	98.8	2799	14	SIVMM316ZX	L22821 Simian immu
11	2609.2	98.8	2781	14	SIVMM132ZO	L22812 Simian immu
12	2609	98.8	10279	14	AY033146	AY033146 Simian im
13	2608.4	98.8	2799	14	SIVMM316ZR	L22815 Simian immu
14	2606	98.7	2781	14	SIVMM132ZL	L22809 Simian immu
15	2603.6	98.6	2799	14	SIVMM316ZQ	L22814 Simian immu
16	2601.2	98.5	2781	14	SIVMM132ZM	L22810 Simian immu
17	2599.6	98.4	2781	14	SIVMM132ZN	L22811 Simian immu
18	2593.2	98.2	2781	14	SIVMM132ZP	L22813 Simian immu
19	2538.6	96.1	10274	14	SIV1A11AA	M76764 Simian immu
20	2533.8	95.9	10277	14	SIVMM32H	D01065 Simian immu
21	2513	95.2	10277	14	SIVMM251	M19499 Simian (mac
22	2509.8	95.0	5391	14	REHTLV4C	X06391 STLV-3 (HTL
23	2476.2	93.8	5111	14	SIU86638	U86638 Simian immu
24	2442.8	92.5	9646	14	RESIVMXX	Y00277 Simian immu
25	2434.6	92.2	7099	6	AX180761	AX180761 Sequence
26	2420.2	91.6	9628	14	SIVMNE	M32741 Simian immu
27	2410.4	91.3	9664	14	SIU79412	U79412 Simian immu
28	2391.2	90.5	2637	14	SIVENVGLYB	L35597 Simian immu
29	2368.8	89.7	2646	14	SIVENVGLYA	L35596 Simian immu
30	2035	77.1	2397	14	SIU95966	U95966 Simian immu
31	1923	72.8	10184	14	AF077017	AF077017 Simian im
32	1877	71.1	9638	14	SIVGAGAB	M80194 Simian immu
33	1873.2	70.9	10246	14	SIV6P12	L09211 Simian immu
34	1873.2	70.9	10246	14	SIV6P6	L09212 Simian immu
35	1873.2	70.9	10246	14	SIV6P9	L09213 Simian immu
36	1868.4	70.7	9996	14	SIVSMMPBJB	L03295 Simian immu
37	1867.2	70.7	9675	14	SIVGAGAA	M80193 Simian immu
38	1866.8	70.7	5094	14	SIVSMMPBJE	L03298 Simian immu
39	1866.8	70.7	9996	14	SIVSMMPBJA	M31325 Simian immu
40	1864.2	70.6	4344	14	SIU04983	U04983 Simian immu
41	1858	70.4	4347	14	SIU04984	U04984 Simian immu
42	1853.6	70.2	4353	14	SIU04990	U04990 Simian immu
43	1853.2	70.2	10241	14	RESIVSMM	X14307 Simian immu
44	1852.2	70.1	5078	14	SIVSMMPBJC	L03296 Simian immu
45	1852	70.1	4352	14	SIU04988	U04988 Simian immu

ALIGNMENTS

RESULT 1
SIVMM239
LOCUS SIVMM239 10535 bp ss-RNA linear VRL 18-APR-2000
DEFINITION Simian (macaque) immunodeficiency virus, isolate 239, complete proviral genome and flanking sequence.
ACCESSION M33262
VERSION M33262.1 GI:334647
KEYWORDS .
SOURCE Simian immunodeficiency virus.
ORGANISM Simian immunodeficiency virus
Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate lentivirus group.
REFERENCE 1 (bases 1 to 10535)
AUTHORS Kestler,H., Kodama,T., Ringler,D., Marthas,M., Pedersen,N.C., Lackner,A., Regier,D., Sehgal,P., Daniel,M., King,N. and

```
            Desrosiers,R.
  JOURNAL   Unpublished
REFERENCE   2  (sites)
  AUTHORS   Kestler,H., Kodama,T., Ringler,D., Marthas,M., Pedersen,N.C.,
            Lackner,A., Regier,D., Sehgal,P., Daniel,M., King,N. and
            Desrosiers,R.
  TITLE     Induction of AIDS in rhesus monkeys by molecularly cloned simian
            immunodeficiency virus
  JOURNAL   Science 248 (4959), 1109-1112 (1990)
  MEDLINE   90260657
   PUBMED   2160735
COMMENT     [2]  sites; for [1].
            Kindly submitted in computer-readable form prior to publication by
            Dr. Desrosiers, New England Regional Primate Center, Harvard.  In
            addition to the 256 bases of flanking cellular DNA shown below,
            approximately 2500 bases of 3' flanking material were determined.
               Author address R.Desrosiers,
               New England Regional Primate Research Center
               Harvard Medical School
               Southborough MA 01772.
FEATURES             Location/Qualifiers
     source          1. .10535
                     /organism="Simian immunodeficiency virus"
                     /isolate="239"
                     /db_xref="taxon:11723"
     LTR             257. .1074
                     /note="5'LTR"
     LTR             257. .773
                     /note="U3"
     repeat_region   774. .950
                     /note="R"
     misc_feature    775
                     /note="mRNA start (putative); putative"
     LTR             951. .1074
                     /note="U5"
     CDS             1309. .2841
                     /note="gag polyprotein"
                     /codon_start=1
                     /protein_id="AAA47632.1"
                     /db_xref="GI:334650"
                     /translation="MGVRNSVLSGKKADELEKIRLRPNGKKKYMLKHVVWAANELDRF
                     GLAESLLENKEGCQKILSVLAPLVPTGSENLKSLYNTVCVIWCIHAEEKVKHTEEAKQ
                     IVQRHLVVETGTTETMPKTSRPTAPSSGRGGNYPVQQIGGNYVHLPLSPRTLNAWVKL
                     IEEKKFGAEVVPGFQALSEGCTPYDINQMLNCVGDHQAAMQIIRDIINEEAADWDLQH
                     PQPAPQQGQLREPSGSDIAGTTSSVDEQIQWMYRQQNPIPVGNIYRRWIQLGLQKCVR
                     MYNPTNILDVKQGPKEPFQSYVDRFYKSLRAEQTDAAVKNWMTQTLLIQNANPDCKLV
                     LKGLGVNPTLEEMLTACQGVGGPGQKARLMAEALKEALAPVPIPFAAAQQRGPRKPIK
                     CWNCGKEGHSARQCRAPRRQGCWKCGKMDHVMAKCPDRQAGFLGLGPWGKKPRNFPMA
                     QVHQGLMPTAPPEDPAVDLLKNYMQLGKQQREKQRESREKPYKEVTEDLLHLNSLFGG
                     DQ"
     CDS             <2484. .5666
                     /note="pol polyprotein"
                     /codon_start=1
                     /protein_id="AAA47633.2"
                     /db_xref="GI:7583329"
                     /translation="VLELWERGTLCKAMQSPKKTGMLEMWKNGPCYGQMPRQTGGFFR
                     PWSMGKEAPQFPHGSSASGADANCSPRGPSCGSAKELHAVGQAAERKAERKQREALQG
                     GDRGFAAPQFSLWRRPVVTAHIEGQPVEVLLDTGADDSIVTGIELGPHYTPKIVGGIG
                     GFINTKEYKNVEIEVLGKRIKGTIMTGDTPINIFGRNLLTALGMSLNFPIAKVEPVKV
                     ALKPGKDGPKLKQWPLSKEKIVALREICEKMEKDGQLEEAPPTNPYNTPTFAIKKKDK
                     NKWRMLIDFRELNRVTQDFTEVQLGIPHPAGLAKRKRITVLDIGDAYFSIPLDEEFRQ
                     YTAFTLPSVNNAEPGKRYIYKVLPQGWKGSPAIFQYTMRHVLEPFRKANPDVTLVQYM
                     DDILIASDRTDLEHDRVVLQSKELLNSIGFSTPEEKFQKDPPFQWMGYELWPTKWKLQ
                     KIELPQRETWTVNDIQKLVGVLNWAAQIYPGIKTKHLCRLIRGKMTLTEEVQWTEMAE
                     AEYEENKIILSQEQEGCYYQEGKPLEATVIKSQDNQWSYKIHQEDKILKVGKFAKIKN
                     THTNGVRLLAHVIQKIGKEAIVIWGQVPKFHLPVEKDVWEQWWTDYWQVTWIPEWDFI
                     STPPLVRLVFNLVKDPIEGEETYYTDGSCNKQSKEGKAGYITDRGKDKVKVLEQTTNQ
                     QAELEAFLMALTDSGPKANIIVDSQYVMGIITGCPTESESRLVNQIIEEMIKKSEIYV
                     AWVPAHKGIGGNQEIDHLVSQGIRQVLFLEKIEPAQEEHDKYHSNVKELVFKFGLPRI
                     VARQIVDTCDKCHQKGEAIHGQANSDLGTWQMDCTHLEGKIIIVAVHVASGFIEAEVI
                     PQETGRQTALFLLKLAGRWPITHLHTDNGANFASQEVKMVAWWAGIEHTFGVPYNPQS
                     QGVVEAMNHHLKNQIDRIREQANSVETIVLMAVHCMNFKRRGGIGDMTPAERLINMIT
                     TEQEIQFQQSKNSKFKNFRVYYREGRDQLWKGPGELLWKGEGAVILKVGTDIKVVPRR
                     KAKIIKDYGGGKEVDSSSHMEDTGEAREVA"
     misc_feature    5340. .5355
                     /note="polypurine tract"
     CDS             5596. .6240
                     /note="vif protein"
                     /codon_start=1
                     /protein_id="AAA47634.1"
                     /db_xref="GI:334652"
                     /translation="MEEEKRWIAVPTWRIPERLERWHSLIKYLKYKTKDLQKVCYVPH
                     FKVGWAWWTCSRVIFPLQEGSHLEVQGYWHLTPEKGWLSTYAVRITWYSKNFWTDVTP
                     NYADILLHSTYFPCFTAGEVRRAIRGEQLLSCCRFPRAHKYQVPSLQYLALKVVSDVR
                     SQGENPTWKQWRRDNRRGLRMAKQNSRGDKQRGGKPPTKGANFPGLAKVLGILA"
     CDS             6068. .6406
                     /note="vpX protein"
                     /codon_start=1
                     /protein_id="AAA47635.1"
                     /db_xref="GI:334653"
                     /translation="MSDPRERIPPGNSGEETIGEAFEWLNRTVEEINREAVNHLPREL
                     IFQVWQRSWEYWHDEQGMSPSYVKYRYLCLIQKALFMHCKKGCRCLGEGHGAGGWRPG
                     PPPPPPPGLA"
     CDS             6407. .6712
                     /note="vpR protein"
                     /codon_start=1
                     /protein_id="AAA47636.1"
                     /db_xref="GI:334654"
                     /translation="MEERPPENEGPQREPWDEWVVEVLEELKEEALKHFDPRLLTALG
                     NHIYNRHGDTLEGAGELIRILQRALFMHFRGGCIHSRIGQPGGGNPLSAIPPSRSML"
     CDS             join(6558. .6853,9062. .9158)
                     /note="tat protein"
                     /codon_start=1
                     /protein_id="AAA47630.1"
                     /db_xref="GI:334648"
                     /translation="METPLREQENSLESSNERSSCISEADASTPESANLGEEILSQLY
                     RPLEACYNTCYCKKCCYHCQFCFLKKGLGICYEQSRKRRRTPKKAKANTSSASNKPIS
                     NRTRHCQPEKAKKETVEKAVATAPGLGR"
     exon            <6558. .6853
                     /note="tat protein,  (first expressed exon)"
                     /number=2
     CDS             join(6784. .6853,9062. .9315)
                     /note="rev protein"
                     /codon_start=1
                     /protein_id="AAA47631.1"
                     /db_xref="GI:334649"
                     /translation="MSNHEREEELRKRLRLIHLLHQTNPYPTGPGTANQRRQRKRRWR
                     RRWQQLLALADRIYSFPDPPTDTPLDLAIQQLQNLAIESIPDPPTNTPEALCDPTEDS
                     RSPQD"
     exon            <6784. .6853
                     /note="rev protein,  (first expressed exon)"
                     /number=2
     CDS             6860. .9499
                     /note="env polyprotein"
                     /codon_start=1
                     /protein_id="AAA47637.1"
                     /db_xref="GI:334655"
                     /translation="MGCLGNQLLIAILLLSVYGIYCTLYVTVFYGVPAWRNATIPLFC
                     ATKNRDTWGTTQCLPDNGDYSEVALNVTESFDAWNNTVTEQAIEDVWQLFETSIKPCV
                     KLSPLCITMRCNKSETDRWGLTKSITTTASTTSTTASAKVDMVNETSSCIAQDNCTGL
                     EQEQMISCKFNMTGLKRDKKKEYNETWYSADLVCEQGNNTGNESRCYMNHCNTSVIQE
                     SCDKHYWDAIRFRYCAPPGYALLRCNDTNYSGFMPKCSKVVVSSCTRMMETQTSTWFG
                     FNGTRAENRTYIYWHGRDNRTIISLNKYYNLTMKCRRPGNKTVLPVTIMSGLVFHSQP
                     INDRPKQAWCWFGGKWKDAIKEVKQTIVKHPRYTGTNNTDKINLTAPGGGDPEVTFMW
                     TNCRGEFLYCKMNWFLNWVEDRNTANQKPKEQHKRNYVPCHIRQIINTWHKVGKNVYL
                     PPREGDLTCNSTVTSLIANIDWIDGNQTNITMSAEVAELYRLELGDYKLVEITPIGLA
                     PTDVKRYTTGGTSRNKRGVFVLGFLGFLATAGSAMGAASLTLTAQSRTLLAGIVQQQQ
                     QLLDVVKRQQELLRLTVWGTKNLQTRVTAIEKYLKDQAQLNAWGCAFRQVCHTTVPWP
                     NASLTPKWNNETWQEWERKVDFLEENITALLEEAQIQQEKNMYELQKLNSWDVFGNWF
                     DLASWIKYIQYGVYIVVGVILLRIVIYIVQMLAKLRQGYRPVFSSPPSYFQQTHIQQD
                     PALPTREGKERDGGEGGGNSSWPWQIEYIHFLIRQLIRLLTWLFSNCRTLLSRVYQIL
                     QPILQRLSATLQRIREVLRTELTYLQYGWSYFHEAVQAVWRSATETLAGAWGDLWETL
                     RRGGRWILAIPRRIRQGLELTLL"
     exon            9062. .>9315
                     /note="rev protein"
                     /number=3
```

```
     exon            9062. .>9158
                     /note="tat protein"
                     /number=3
     CDS             9333. .10124
                     /note="nef protein (in-frame stop at 9612)"
                     /pseudo
                     /codon_start=1
     CDS             9333. .9611
                     /note="nef protein (premature termination)"
                     /codon_start=1
                     /protein_id="AAA47638.1"
                     /db_xref="GI:334656"
                     /translation="MGGAISMRRSRPSGDLRQRLLRARGETYGRLLGEVEDGYSQSPG
                     GLDKGLSSLSCEGQKYNQGQYMNTPWRNPAEEREKLAYRKQNMDDIDE"
     misc_feature    9701. .9716
                     /note="polypurine tract"
     LTR             9718. .10535
                     /note="3' LTR"
     LTR             9718. .10234
                     /note="U3"
     repeat_region   10235. .10411
                     /note="R"
     polyA_signal    10388. .10393
     LTR             10412. .10535
                     /note="U5"
BASE COUNT     3540 a   1988 c   2607 g   2400 t
ORIGIN

  Query Match             100.0%;  Score 2641;  DB 14;  Length 10535;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2641;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TCTAGAAGCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6695 TCTAGAAGCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTG 6754

Qy      61 TTTTCTTAAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6755 TTTTCTTAAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCC 6814

Qy     121 GAAAAAGGCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6815 GAAAAAGGCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGG 6874

Qy     181 AATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6875 AATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATAT 6934

Qy     241 GTCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6935 GTCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCA 6994

Qy     301 ACCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6995 ACCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCA 7054

Qy     361 GAAGTGGCCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7055 GAAGTGGCCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAG 7114

Qy     421 GCAATAGAGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7115 GCAATAGAGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCC 7174

Qy     481 CCATTATGCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7175 CCATTATGCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAA 7234

Qy     541 TCAATAACAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7235 TCAATAACAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTC 7294

Qy     601 AATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCAAATG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7295 AATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCAAATG 7354

Qy     661 ATAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAATGAA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7355 ATAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAATGAA 7414

Qy     721 ACTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7415 ACTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGA 7474

Qy     781 TGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7475 TGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGG 7534

Qy     841 GATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGAC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7535 GATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGAC 7594

Qy     901 ACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7595 ACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGG 7654

Qy     961 ATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGA 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7655 ATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGA 7714

Qy    1021 ACTTATATTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTAT 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7715 ACTTATATTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTAT 7774

Qy    1081 AATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATG 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7775 AATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATG 7834

Qy    1141 TCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGGCCAAAGCAGGCATGGTGTTGG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7835 TCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGGCCAAAGCAGGCATGGTGTTGG 7894

Qy    1201 TTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCC 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7895 TTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCC 7954

Qy    1261 AGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGAGGAGGAGAT 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7955 AGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGAGGAGGAGAT 8014

Qy    1321 CCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATGAAT 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8015 CCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATGAAT 8074

Qy    1381 TGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGGAACAGCAT 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8075 TGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGGAACAGCAT 8134

Qy    1441 AAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAGGC 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8135 AAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAGGC 8194

Qy    1501 AAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACCAGT 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8195 AAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACCAGT 8254

Qy    1561 CTCATAGCAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTGCAGAG 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8255 CTCATAGCAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTGCAGAG 8314

Qy    1621 GTGGCAGAACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTCCAATT 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8315 GTGGCAGAACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTCCAATT 8374

Qy    1681 GGCTTGGCCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAAGA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    8375 GGCTTGGCCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAAGA 8434

Qy    1741 GGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCG 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8435 GGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCG 8494

Qy    1801 GCGTCGTTGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGCAACAG 1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8495 GCGTCGTTGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGCAACAG 8554

Qy    1861 CAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCTGGGGA 1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8555 CAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCTGGGGA 8614

Qy    1921 ACAAAGAACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAG 1980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8615 ACAAAGAACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAG 8674

Qy    1981 CTGAATGCTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGCCAAAT 2040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8675 CTGAATGCTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGCCAAAT 8734

Qy    2041 GCAAGTCTAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGAC 2100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8735 GCAAGTCTAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGAC 8794

Qy    2101 TTCTTGGAAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAAC 2160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8795 TTCTTGGAAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAAC 8854

Qy    2161 ATGTATGAATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCT 2220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8855 ATGTATGAATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCT 8914

Qy    2221 TCTTGGATAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGA 2280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8915 TCTTGGATAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGA 8974

Qy    2281 ATAGTGATCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTC 2340
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8975 ATAGTGATCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTC 9034

Qy    2341 TCTTCCCCACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACC 2400
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9035 TCTTCCCCACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACC 9094

Qy    2401 AGAGAAGGCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAG 2460
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9095 AGAGAAGGCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAG 9154

Qy    2461 ATAGAATATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGC 2520
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9155 ATAGAATATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGC 9214

Qy    2521 AACTGCAGAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTC 2580
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9215 AACTGCAGAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTC 9274

Qy    2581 TCTGCGACCCTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTACCTACAATAT 2640
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9275 TCTGCGACCCTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTACCTACAATAT 9334

Qy    2641 G 2641
           |
Db    9335 G 9335


RESULT 2
SIVMM316ZZ
LOCUS       SIVMM316ZZ              2799 bp ss-DNA     linear   VRL 21-DEC-1993
DEFINITION  Simian immunodeficiency virus (Mm316zm) envelope glycoprotein and
            nef protein genes, complete cds and 5'end.
ACCESSION   L22823
```

```
VERSION     L22823.1  GI:431599
KEYWORDS    envelope glycoprotein; nef protein.
SOURCE      Simian immunodeficiency virus DNA.
  ORGANISM  Simian immunodeficiency virus
            Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
            lentivirus group.
REFERENCE   1  (bases 1 to 2799)
  AUTHORS   Kodama,T., Mori,K., Kawahara,T., Ringler,D.J. and Desrosiers,R.C.
  TITLE     Analysis of simian immunodeficiency virus sequence variation in
            tissues of rhesus macaques with simian AIDS
  JOURNAL   J. Virol. 67 (11), 6522-6534 (1993)
  MEDLINE   94016835
   PUBMED   8411355
FEATURES             Location/Qualifiers
     source          1. .2799
                     /organism="Simian immunodeficiency virus"
                     /proviral
                     /db_xref="taxon:11723"
     gene            1. .2457
                     /gene="tat"
     exon            1. .150
                     /gene="tat"
                     /number=1
     gene            83. .2614
                     /gene="rev"
     exon            83. .150
                     /gene="tat"
                     /number=1
     misc_signal     150
                     /gene="tat"
                     /note="splice donor site"
     gene            159. .2798
                     /gene="env"
     CDS             159. .2798
                     /gene="env"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAA47708.1"
                     /db_xref="GI:431600"
                     /translation="MGCLGNQLLIAILLLSVYGIYCTLYVTVFYGVPAWRNATIPLFC
                     ATKNRDTWGTTQCLPDNGDYSEMALNVTESFDAWNNTVTEQAIEDVWQLFETSIKPCV
                     KLSPLCITMRCNKSETDRWGLTKSITTTASTTSTTASAKVDMVNETSSCIAQDNCAGL
                     EQEQMISCKFNMTGLKRNKKKEYNETWYSADLVCEQGNNTGNESRCYMNHCNTSVIQE
                     SCDKHYWDAIRFRYCAPPGYALLRCNDTNYSGFMPKCSKVVVSSCTRMMETQTSTWFG
                     FNGTRAENRTYIYWHGRDNRTIISLNKYYNLTMKCRRPGNKTVLPVTIMSGLVFHSQL
                     INDRPKQAWCWFGGKWKDAIKEVKQTIVKHPRYTGTSNTDKINLTAPRGGNPEVTFMW
                     TNCRGEFLYCKMNWFLNWVEDRNTANQKPKEQHKRNYVPCHIRQIINTWHKVGKNVYL
                     PPREGDLTCNSTVTSLIANIDWIDGNQTNITMSAEVAELYRLELGDYKLVEITPIGLA
                     PTDVKRYTTGGTSRNKRGVFVLGFLGFLATAGSAMGAASLTLTAQSRTLLAGIVQQQQ
                     QLLDVVKRQQELLRLTVWGTKNLQTRVTAIEKYLKDQAQLNAWGCAFRQVCHTTVPWP
                     NASLTPKWNNETWQEWERKVDFLEENITALLEEAQIQQEKNMYELQKLNSWDVFGNWF
                     DLASWIKYIQYGVYIVVGVILLRIVIYIVQMLAKLRQGYRPVFSSPPSYFQQTHIQQD
                     PALPTREGKEGDGGEGGGNSSWPWQIEYIHFLIRQLIRLLTWLFSNCRTLLSRVYQIL
                     QPILQRLSATLQRIREVLRTELTYLQYGWSYFHEAVQAVWRSAAETLAGAWGDLWETL
                     RRGGRWILAIPRRIRQGLELTLL"
     misc_signal     1734
                     /gene="tat"
                     /note="SU/TM cleavage site"
     exon            2358. .2614
                     /gene="rev"
                     /number=2
     exon            2358. .2457
                     /gene="tat"
                     /number=2
     misc_signal     2358
                     /gene="tat"
                     /note="splice acceptor"
     gene            2632. .2799
                     /gene="nef"
     CDS             2632. .>2799
                     /gene="nef"
                     /codon_start=1
                     /protein_id="AAA47709.1"
```

```
                 /db_xref="GI:431601"
                 /translation="MGGAISMRRSRPSGDLRQRLLRARGETYGRLLGEVEDGYSQSPG
                 GLDKGLSSLSCE"
BASE COUNT      900 a    525 c    671 g    703 t
ORIGIN

  Query Match           99.1%;  Score 2618;  DB 14;  Length 2799;
  Best Local Similarity   99.6%;  Pred. No. 0;
  Matches 2624;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       8 GCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTGTTTTCTT 67
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTGTTTTCTT 60

Qy      68 AAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCCGAAAAAG 127
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCCGAAAAAG 120

Qy     128 GCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGGAATCAGC 187
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGGAATCAGC 180

Qy     188 TGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATATGTCACAG 247
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATATGTCACAG 240

Qy     248 TCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCAACCAAGA 307
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     241 TCTTTTATGGTGTGCCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCAACCAAGA 300

Qy     308 ATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCAGAAGTGG 367
          ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db     301 ATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCAGAAATGG 360

Qy     368 CCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAGGCAATAG 427
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAGGCAATAG 420

Qy     428 AGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCCCCATTAT 487
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 AGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCCCCATTAT 480

Qy     488 GCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAATCAATAA 547
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 GCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAATCAATAA 540

Qy     548 CAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTCAATGAGA 607
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 CAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTCAATGAGA 600

Qy     608 CTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCAAATGATAAGCT 667
          |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db     601 CTAGTTCTTGTATAGCCCAGGATAATTGCGCAGGCTTGGAACAAGAGCAAATGATAAGCT 660

Qy     668 GTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAATGAAACTTGGT 727
          |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db     661 GTAAATTCAACATGACAGGGTTAAAAAGAAACAAGAAAAAAGAGTACAATGAAACTTGGT 720

Qy     728 ACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGATGTTACA 787
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 ACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGATGTTACA 780

Qy     788 TGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGGGATGCTA 847
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 TGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGGGATGCTA 840

Qy     848 TTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGACACAAATT 907
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 TTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGACACAAATT 900

Qy     908 ATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGGATGATGG 967
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 ATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGGATGATGG 960

Qy     968 AGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGAACTTATA 1027
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 AGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGAACTTATA 1020

Qy    1028 TTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTATAATCTAA 1087
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 TTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTATAATCTAA 1080

Qy    1088 CAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATGTCTGGAT 1147
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 CAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATGTCTGGAT 1140

Qy    1148 TGGTTTTCCACTCACAACCAATCAATGATAGGCCAAAGCAGGCATGGTGTTGGTTTGGAG 1207
          ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    1141 TGGTTTTCCACTCACAACTAATCAATGATAGGCCAAAGCAGGCATGGTGTTGGTTTGGAG 1200

Qy    1208 GAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCCAGGTATA 1267
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 GAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCCAGGTATA 1260

Qy    1268 CTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGAGGAGGAGATCCGGAAG 1327
          |||||||| |||||||||||||||||||||||||||||| |||||||| ||||||||||||
Db    1261 CTGGAACTAGCAATACTGATAAAATCAATTTGACGGCTCCTAGAGGAGGAAATCCGGAAG 1320

Qy    1328 TTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATGAATTGGTTTC 1387
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 TTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATGAATTGGTTTC 1380

Qy    1388 TAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGGAACAGCATAAAAGGA 1447
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db    1381 TAAATTGGGTAGAAGATAGGAATACAGCTAACCAAAAGCCAAAGGAACAGCATAAAAGGA 1440

Qy    1448 ATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAGGCAAAAATG 1507
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 ATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAGGCAAAAATG 1500

Qy    1508 TTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACCAGTCTCATAG 1567
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 TTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACCAGTCTCATAG 1560

Qy    1568 CAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTGCAGAGGTGGCAG 1627
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 CAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTGCAGAGGTGGCAG 1620

Qy    1628 AACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTCCAATTGGCTTGG 1687
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 AACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTCCAATTGGCTTGG 1680

Qy    1688 CCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCT 1747
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1681 CCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCT 1740

Qy    1748 TTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGT 1807
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1741 TTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGT 1800

Qy    1808 TGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGC 1867
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1801 TGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGC 1860

Qy    1868 TGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACAAAGA 1927
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1861 TGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACAAAGA 1920

Qy    1928 ACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATG 1987
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1921 ACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATG 1980

Qy    1988 CTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGCCAAATGCAAGTC 2047
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1981 CTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGCCAAATGCAAGTC 2040
```

```
Qy    2048 TAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGG 2107
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2041 TAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGG 2100

Qy    2108 AAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATG 2167
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2101 AAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATG 2160

Qy    2168 AATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGA 2227
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2161 AATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGA 2220

Qy    2228 TAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGA 2287
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2221 TAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGA 2280

Qy    2288 TCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTCTCTTCCC 2347
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2281 TCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTCTCTTCCC 2340

Qy    2348 CACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACCAGAGAAG 2407
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2341 CACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACCAGAGAAG 2400

Qy    2408 GCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAAT 2467
           |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2401 GCAAAGAAGGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAAT 2460

Qy    2468 ATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGCAACTGCA 2527
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2461 ATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGCAACTGCA 2520

Qy    2528 GAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTCTCTGCGA 2587
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2521 GAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTCTCTGCGA 2580

Qy    2588 CCCTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTACCTACAATATG 2641
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2581 CCCTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTACCTACAATATG 2634


RESULT 3
SIVMM316ZW
LOCUS       SIVMM316ZW              2799 bp ss-DNA     linear   VRL 21-DEC-1993
DEFINITION  Simian immunodeficiency virus (Mm361zk) envelope glycoprotein and
            nef protein genes, complete cds and 5'end.
ACCESSION   L22820
VERSION     L22820.1  GI:431590
KEYWORDS    envelope glycoprotein; nef protein.
SOURCE      Simian immunodeficiency virus DNA.
  ORGANISM  Simian immunodeficiency virus
            Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
            lentivirus group.
REFERENCE   1  (bases 1 to 2799)
  AUTHORS   Kodama,T., Mori,K., Kawahara,T., Ringler,D.J. and Desrosiers,R.C.
  TITLE     Analysis of simian immunodeficiency virus sequence variation in
            tissues of rhesus macaques with simian AIDS
  JOURNAL   J. Virol. 67 (11), 6522-6534 (1993)
  MEDLINE   94016835
   PUBMED   8411355
FEATURES             Location/Qualifiers
     source          1. .2799
                     /organism="Simian immunodeficiency virus"
                     /proviral
                     /db_xref="taxon:11723"
     gene            1. .150
                     /gene="tat"
     exon            1. .150
                     /gene="tat"
                     /number=1
     gene            83. .150
                     /gene="rev"
     exon            83. .150
                     /gene="rev"
                     /number=1
     misc_signal     150
                     /gene="rev"
                     /note="splice donor site"
     CDS             159. .2798
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAA47702.1"
                     /db_xref="GI:431591"
                     /translation="MGCLGNQLLIAILLLSVYGIYCTLYVTVFYGVPAWRNATIPLFC
                     ATKNRDTWGTTQCLPDNGDYSEMALNVIESFDAWNNTVTEQAIEDVWQLFETSIKPCV
                     KLSPLCITMRCNKSETDRWGLTKSITTTASTSTTASAKVDMVNETSSCIAQDNCTGL
                     EQEQMISCKFNMTGLKRNKKKEYNETWYSADLVCEQGNNTGNESRCYMNHCNTSVIQE
                     SCDKHYWDAIRFRYCAPPGYALLRCNDTNYSGFMPKCAKVVVSSCTRMMETQTSTWFG
                     FNGTRAENRTYIYWHGRDNRTIISLNKYYNLTMKCRRPGNKTVLPVAIMSGLVFHSQL
                     INDRPKQAWCWFGGKWKDAIKEVKQTIVKHPRYTGTNNTDKINLTAPRGGNPEVTFMW
                     TNCRGEFLYCKMNWFLNWVEDRNTANQKSKEQHKRNYVPCHIRQIINTWHKVGKNVYL
                     PPREGDLTCNSTVTSLIANIDWIDGNQTNITMSAEVAELYRLELGDYKLVEITPIGLA
                     PTDVKRYTTGGTSRNKRGVFVLGFLGFLATAGSAMGAASLTLTAQSRTLLAGIVQQQQ
                     QLLDVVKRQQELLRLTVWGTKNLQTRVTAIEKYLKDQAQLNAWGCAFRQVCHTTVPWP
                     NASLTPKWNNETWQEWERKVDFLEENITALLEEAQIQQEKNMYELQKLNSWDVFGNWF
                     DLASWIKYIQYGVYIVVGVILLRIVIYIVQMLAKLRQGYRPVFSSPPSYFQQTHIQQD
                     PALPTREGKEGDGGEGGGNSSWPWQIEYIHFLIRQLIRLLTWLFSNCRTLLSRVYQIL
                     QPILQRLSATLQRIREVLRTELTYLQYGWSYFHEAVQAVWRSATETLAGAWGDLWETL
                     RRGGRWILAIPRRIRQGLELTLL"
     misc_signal     1734
                     /note="SU/TM cleavage site"
     exon            2358. .2614
                     /product="Rev (exon 2)"
     exon            2358. .2457
                     /product="Tat (exon 2)"
     misc_signal     2358
                     /note="splice acceptor"
     gene            2632. .2799
                     /gene="nef"
     CDS             2632. .>2799
                     /gene="nef"
                     /codon_start=1
                     /product="nef protein"
                     /protein_id="AAA47703.1"
                     /db_xref="GI:431592"
                     /translation="MGGAISMRRSRPSGDLRQRLLRARGETYGRLLGEVEDGYSQSPG
                     GLDKGLSSLSCE"
BASE COUNT      902 a    523 c    670 g    704 t
ORIGIN

  Query Match           99.0%;  Score 2614.8;  DB 14;  Length 2799;
  Best Local Similarity   99.5%;  Pred. No. 0;
  Matches 2622;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       8 GCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTGTTTTCTT 67
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTGTTTTCTT 60

Qy      68 AAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCCGAAAAAG 127
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCCGAAAAAG 120

Qy     128 GCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGGAATCAGC 187
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGGAATCAGC 180

Qy     188 TGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATATGTCACAG 247
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATATGTCACAG 240

Qy     248 TCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCAACCAAGA 307
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCAACCAAGA 300

Qy     308 ATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCAGAAGTGG 367
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
```

```
Db    301 ATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCAGAAATGG 360

Qy    368 CCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAGGCAATAG 427
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CCCTTAATGTTATAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAGGCAATAG 420

Qy    428 AGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCCCCATTAT 487
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCCCCATTAT 480

Qy    488 GCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAATCAATAA 547
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAATCAATAA 540

Qy    548 CAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTCAATGAGA 607
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTCAATGAGA 600

Qy    608 CTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCAAATGATAAGCT 667
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCAAATGATAAGCT 660

Qy    668 GTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAATGAAACTTGGT 727
          |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db    661 GTAAATTCAACATGACAGGGTTAAAAAGAAACAAGAAAAAAGAGTACAATGAAACTTGGT 720

Qy    728 ACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGATGTTACA 787
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 ACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGATGTTACA 780

Qy    788 TGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGGGATGCTA 847
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 TGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGGGATGCTA 840

Qy    848 TTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGACACAAATT 907
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 TTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGACACAAATT 900

Qy    908 ATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGGATGATGG 967
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    901 ATTCAGGCTTTATGCCTAAATGTGCTAAGGTGGTGGTCTCTTCATGCACAAGGATGATGG 960

Qy    968 AGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGAACTTATA 1027
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 AGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGAACTTATA 1020

Qy   1028 TTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTATAATCTAA 1087
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 TTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTATAATCTAA 1080

Qy   1088 CAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATGTCTGGAT 1147
          |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   1081 CAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCGCCATTATGTCTGGAT 1140

Qy   1148 TGGTTTTCCACTCACAACCAATCAATGATAGGCCAAAGCAGGCATGGTGTTGGTTTGGAG 1207
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 TGGTTTTCCACTCACAACTAATCAATGATAGGCCAAAGCAGGCATGGTGTTGGTTTGGAG 1200

Qy   1208 GAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCCAGGTATA 1267
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 GAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCCAGGTATA 1260

Qy   1268 CTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGAGGAGGAGATCCGGAAG 1327
          |||||||||||||||||||||||||||||||||||||||| |||||||| |||||||||
Db   1261 CTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTAGAGGAGGAAATCCGGAAG 1320

Qy   1328 TTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATGAATTGGTTTC 1387
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 TTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATGAATTGGTTTC 1380

Qy   1388 TAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGGAACAGCATAAAAGGA 1447
          |||||||||||||||||||||||||||||||| ||| ||||||||||||||||||||||
Db   1381 TAAATTGGGTAGAAGATAGGAATACAGCTAACCAAAAGTCAAAGGAACAGCATAAAAGGA 1440

Qy   1448 ATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAGGCAAAAATG 1507
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 ATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAGGCAAAAATG 1500

Qy   1508 TTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACCAGTCTCATAG 1567
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 TTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACCAGTCTCATAG 1560

Qy   1568 CAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTGCAGAGGTGGCAG 1627
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 CAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTGCAGAGGTGGCAG 1620

Qy   1628 AACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTCCAATTGGCTTGG 1687
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 AACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTCCAATTGGCTTGG 1680

Qy   1688 CCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCT 1747
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 CCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCT 1740

Qy   1748 TTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGT 1807
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 TTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGT 1800

Qy   1808 TGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGC 1867
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 TGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGC 1860

Qy   1868 TGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACAAAGA 1927
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 TGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACAAAGA 1920

Qy   1928 ACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATG 1987
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 ACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATG 1980

Qy   1988 CTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGCCAAATGCAAGTC 2047
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 CTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGCCAAATGCAAGTC 2040

Qy   2048 TAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGG 2107
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 TAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGG 2100

Qy   2108 AAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATG 2167
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 AGGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATG 2160

Qy   2168 AATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGA 2227
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 AATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGA 2220

Qy   2228 TAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGA 2287
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 TAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGA 2280

Qy   2288 TCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTCTCTTCCC 2347
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 TCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTCTCTTCCC 2340

Qy   2348 CACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACCAGAGAAG 2407
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 CACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACCAGAGAAG 2400

Qy   2408 GCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAAT 2467
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 GCAAAGAAGGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAAT 2460

Qy   2468 ATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGCAACTGCA 2527
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 ATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGCAACTGCA 2520
```

```
Qy    2528 GAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTCTCTGCGA 2587
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2521 GAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTCTCTGCGA 2580

Qy    2588 CCCTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTACCTACAATATG 2641
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2581 CCCTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTACCTACAATATG 2634


RESULT 4
SIVMM316ZS
LOCUS       SIVMM316ZS              2799 bp ss-DNA     linear   VRL 21-DEC-1993
DEFINITION  Simian immunodeficiency virus (Mm316zg) envelope glycoprotein and
            nef protein genes, complete cds and 5'end.
ACCESSION   L22816
VERSION     L22816.1  GI:431579
KEYWORDS    envelope glycoprotein; nef protein.
SOURCE      Simian immunodeficiency virus DNA.
  ORGANISM  Simian immunodeficiency virus
            Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
            lentivirus group.
REFERENCE   1  (bases 1 to 2799)
  AUTHORS   Kodama,T., Mori,K., Kawahara,T., Ringler,D.J. and Desrosiers,R.C.
  TITLE     Analysis of simian immunodeficiency virus sequence variation in
            tissues of rhesus macaques with simian AIDS
  JOURNAL   J. Virol. 67 (11), 6522-6534 (1993)
  MEDLINE   94016835
   PUBMED   8411355
FEATURES             Location/Qualifiers
     source          1. .2799
                     /organism="Simian immunodeficiency virus"
                     /proviral
                     /db_xref="taxon:11723"
     gene            1. .2457
                     /gene="tat"
     exon            1. .150
                     /gene="tat"
                     /number=1
     gene            83. .2614
                     /gene="rev"
     exon            83. .150
                     /gene="tat"
                     /number=1
     misc_signal     150
                     /gene="tat"
                     /note="splice donor site"
     gene            159. .2798
                     /gene="env"
     CDS             159. .2798
                     /gene="env"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAA47696.1"
                     /db_xref="GI:431580"
                     /translation="MGCLGNQLLIAILFLSVYGIYCTLYVTVFYGVPAWRNATIPLFC
                     ATKNRDTWGTTQCLPDNGDYSEMALNVTESFDAWNNTVTEQAIEDVWQLFETSIKPCV
                     KLSPLCITMRCNKSETDRWGLTKSITTTASTTSTTASAKVDMVNETSSCIAQDNCTGL
                     EQEQMISCKFNMTGLKRDKKKEYNETWYSADLVCEQGNNTGNESRCYMNHCNTSVIQE
                     SCDKHYWDAIRFRYCAPPGYALLRCNDTNYSGFMPKCSKVVVSSCTRMMETQTSTWFG
                     FNGTRAENRTYIYWHGRDNRTIISLNKYYNLTMKCRRPGNKTVLPVTIMSGLVFHSQP
                     INNRPKQAWCWFGGKWKDAIKEVKQTIVKHPRYTGTNNTDKINLTAPRGGNPEVTFMW
                     TNCRGEFLYCKMNWFLNWVEDRNTANQKSKEQHKRNYVPCHIRQIINTWYKVGKNVYL
                     PPREGDLTCNSTVTSLIANIDWIDGNQTNITMSAEVAELYRLELGDYKLVEITPIGLA
                     PTDVKRYTTGGTSRNKRGVFVLGFLGFLATAGSAMGAASLTLTAQSRTLLAGIVQQQQ
                     QLLDVVTRQQELLRLTVWGTKNLQTRVTAIEKYLKDQAQLNAWGCAFRQVCHTTVPWP
                     NASLTPKWNNETWQEWERKVDFLEENITALLEEAQIQQEKNMYELQKLNSWDVFGNWF
                     DLASWIKYIQYGVYIVVGVILLRIVIYIVQMLAELRQGYRPVFSSPPSYFQQTHIQQD
                     PALPTREGKERDGGEGGGNSSWPWQIEYIHFLIRQLIRLLTWLFSNCRTLLSRVYQIL
                     QPILQRLSATLQRIREVLRTELTYLQYGWSYFHEAVQAVWRSATETLAGAWGDLWETL
                     RRGGRWILAIPRRIRQGLELTLL"
     misc_signal     1734
                     /gene="tat"
                     /note="SU/TM cleavage site"
     exon            2358. .2614
                     /gene="rev"
                     /number=2
     exon            2358. .2457
                     /gene="tat"
                     /number=2
     misc_signal     2358
                     /gene="tat"
                     /note="splice acceptor"
     gene            2632. .2799
                     /gene="nef"
     CDS             2632. .>2799
                     /gene="nef"
                     /codon_start=1
                     /product="nef protein"
                     /protein_id="AAA47697.1"
                     /db_xref="GI:431581"
                     /translation="MGGAISMRRSRPSGDLRQRLLRARGETYGRLLGEVEDGYSQSPG
                     GLDKGLSSLSCE"
BASE COUNT      902 a    525 c    667 g    705 t
ORIGIN

  Query Match             98.9%;  Score 2613.2;  DB 14;  Length 2799;
  Best Local Similarity   99.5%;  Pred. No. 0;
  Matches 2621;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy      8 GCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTGTTTTCTT 67
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTGTTTTCTT 60

Qy     68 AAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCCGAAAAAG 127
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     61 AAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGCAGAACTCCGAAAAAG 120

Qy    128 GCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGGAATCAGC 187
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGGAATCAGC 180

Qy    188 TGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATATGTCACAG 247
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    181 TGCTTATCGCCATCTTGTTTTAAGTGTCTATGGGATCTATTGTACTCTATATGTCACAG 240

Qy    248 TCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCAACCAAGA 307
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCAACCAAGA 300

Qy    308 ATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCAGAAGTGG 367
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    301 ATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCAGAAATGG 360

Qy    368 CCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAGGCAATAG 427
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAGGCAATAG 420

Qy    428 AGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCCCCATTAT 487
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCCCCATTAT 480

Qy    488 GCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAATCAATAA 547
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAATCAATAA 540

Qy    548 CAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTCAATGAGA 607
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTCAATGAGA 600

Qy    608 CTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCAAATGATAAGCT 667
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCAAATGATAAGCT 660

Qy    668 GTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAATGAAACTTGGT 727
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAATGAAACTTGGT 720
```

```
Qy    728 ACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGATGTTACA 787
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 ACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGATGTTACA 780

Qy    788 TGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGGGATGCTA 847
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 TGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGGGATGCTA 840

Qy    848 TTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGACACAAATT 907
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 TTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGACACAAATT 900

Qy    908 ATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGGATGATGG 967
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 ATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGGATGATGG 960

Qy    968 AGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGAACTTATA 1027
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 AGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGAACTTATA 1020

Qy   1028 TTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTATAATCTAA 1087
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 TTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTATAATCTAA 1080

Qy   1088 CAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATGTCTGGAT 1147
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   1081 CAATGAAATGTAGAAGACCAGGAAATAAGACAGTCTTACCAGTCACCATTATGTCTGGAT 1140

Qy   1148 TGGTTTTCCACTCACAACCAATCAATGATAGGCCAAAGCAGGCATGGTGTTGGTTTGGAG 1207
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   1141 TGGTTTTCCACTCACAACCAATCAATAATAGGCCAAAGCAGGCATGGTGTTGGTTTGGAG 1200

Qy   1208 GAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCCAGGTATA 1267
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 GAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCCAGGTATA 1260

Qy   1268 CTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGAGGAGGAGATCCGGAAG 1327
          |||||||||||||||||||||||||||||||||||||||| |||||||| |||||||||
Db   1261 CTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTAGAGGAGGAAATCCGGAAG 1320

Qy   1328 TTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATGAATTGGTTTC 1387
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 TTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATGAATTGGTTTC 1380

Qy   1388 TAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGGAACAGCATAAAAGGA 1447
          |||||||||||||||||||||||||||||||| ||| ||||||||||||||||||||||||
Db   1381 TAAATTGGGTAGAAGATAGGAATACAGCTAACCAAAAGTCAAAGGAACAGCATAAAAGGA 1440

Qy   1448 ATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAGGCAAAAATG 1507
          |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db   1441 ATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGTATAAAGTAGGCAAAAATG 1500

Qy   1508 TTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACCAGTCTCATAG 1567
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 TTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACCAGTCTCATAG 1560

Qy   1568 CAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTGCAGAGGTGGCAG 1627
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 CAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTGCAGAGGTGGCAG 1620

Qy   1628 AACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTCCAATTGGCTTGG 1687
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 AACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTCCAATTGGCTTGG 1680

Qy   1688 CCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCT 1747
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 CCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCT 1740

Qy   1748 TTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGT 1807
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 TTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGT 1800
```

```
Qy   1808 TGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGC 1867
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 TGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGC 1860

Qy   1868 TGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACAAAGA 1927
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 TGTTGGACGTGGTCACGAGACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACAAAGA 1920

Qy   1928 ACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATG 1987
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 ACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATG 1980

Qy   1988 CTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGCCAAATGCAAGTC 2047
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 CTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGCCAAATGCAAGTC 2040

Qy   2048 TAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGG 2107
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 TAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGG 2100

Qy   2108 AAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATG 2167
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 AAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATG 2160

Qy   2168 AATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGA 2227
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 AATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGA 2220

Qy   2228 TAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGA 2287
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 TAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGA 2280

Qy   2288 TCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTCTCTTCCC 2347
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   2281 TCTATATAGTACAAATGCTAGCTGAGTTAAGGCAGGGGTATAGGCCAGTGTTCTCTTCCC 2340

Qy   2348 CACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACCAGAGAAG 2407
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 CACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACCAGAGAAG 2400

Qy   2408 GCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAAT 2467
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 GCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAAT 2460

Qy   2468 ATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGCAACTGCA 2527
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 ATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGCAACTGCA 2520

Qy   2528 GAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTCTCTGCGA 2587
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 GAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTCTCTGCGA 2580

Qy   2588 CCCTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTACCTACAATATG 2641
          || ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 CCTTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTACCTACAATATG 2634
```

RESULT 5
SIVMM316ZU

LOCUS	SIVMM316ZU 2799 bp ss-DNA linear VRL 13-OCT-1995
DEFINITION	Simian immunodeficiency virus (Mm316zi) envelope glycoprotein and nef protein genes, complete cds and 5'end.
ACCESSION	L22818
VERSION	L22818.1 GI:431585
KEYWORDS	envelope glycoprotein; nef protein.
SOURCE	Simian immunodeficiency virus DNA.
ORGANISM	Simian immunodeficiency virus Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate lentivirus group.
REFERENCE	1 (bases 1 to 2799)
AUTHORS	Kodama,T., Mori,K., Kawahara,T., Ringler,D.J. and Desrosiers,R.C.
TITLE	Analysis of simian immunodeficiency virus sequence variation in tissues of rhesus macaques with simian AIDS

```
  JOURNAL   J. Virol. 67 (11), 6522-6534 (1993)
  MEDLINE   94016835
   PUBMED   8411355
FEATURES             Location/Qualifiers
     source          1. .2799
                     /organism="Simian immunodeficiency virus"
                     /proviral
                     /db_xref="taxon:11723"
     gene            1. .2457
                     /gene="tat"
     exon            1. .150
                     /gene="tat"
                     /number=1
     gene            83. .2614
                     /gene="rev"
     exon            83. .150
                     /gene="rev"
                     /number=1
     misc_signal     150
                     /gene="tat"
                     /note="splice donor site"
     gene            159. .2798
                     /gene="env"
     CDS             159. .2798
                     /gene="env"
                     /note="internal stop codon at 519 (TAG) and art 1386
                     (TAG)"
                     /pseudo
                     /codon_start=1
                     /product="envelope glycoprotein"
     misc_signal     1734
                     /gene="tat"
                     /note="SU/TM cleavage site"
     exon            2358. .2614
                     /gene="rev"
                     /number=2
     exon            2358. .2457
                     /gene="tat"
                     /number=2
     misc_signal     2358
                     /gene="tat"
                     /note="splice acceptor"
     gene            2632. .2799
                     /gene="nef"
     CDS             2632. .>2799
                     /gene="nef"
                     /codon_start=1
                     /product="nef protein"
                     /protein_id="AAA79276.1"
                     /db_xref="GI:431586"
                     /translation="MGGAISMRRSRPSGDLRQRLLRARGETYGRLLGEVEDGYSQSPG
                     GLDKGLSSLSCE"
BASE COUNT      904 a    524 c    668 g    703 t
ORIGIN

  Query Match             98.9%;  Score 2613.2;  DB 14;  Length 2799;
  Best Local Similarity   99.5%;  Pred. No. 0;
  Matches 2621;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy       8 GCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTGTTTTCTT 67
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTGTTTTCTT 60

Qy      68 AAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCCGAAAAAG 127
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCCGAAAAAG 120

Qy     128 GCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGGAATCAGC 187
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGGAATCAGC 180

Qy     188 TGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATATGTCACAG 247
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db     181 TGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATATGTCACAG 240

Qy     248 TCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCAACCAAGA 307
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCAACCAAGA 300

Qy     308 ATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCAGAAGTGG 367
           ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db     301 ATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCAGAAATGG 360

Qy     368 CCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAGGCAATAG 427
           ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db     361 CCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACGGGCAATAG 420

Qy     428 AGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCCCCATTAT 487
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 AGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCCCCATTAT 480

Qy     488 GCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAATCAATAA 547
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     481 GCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATAGGGATTGACAAAATCAATAA 540

Qy     548 CAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTCAATGAGA 607
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 CAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTCAATGAGA 600

Qy     608 CTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCAAATGATAAGCT 667
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 CTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCAAATGATAAGCT 660

Qy     668 GTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAATGAAACTTGGT 727
           |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db     661 GTAAATTCAACATGACAGGGTTAAAAAGAAACAAGAAAAAAGAGTACAATGAAACTTGGT 720

Qy     728 ACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGATGTTACA 787
           |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db     721 ACTCTGCAGATTTGGTATGTGAACAGGGGAATAACACTGGTAATGAAAGTAGATGTTACA 780

Qy     788 TGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGGGATGCTA 847
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 TGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGGGATGCTA 840

Qy     848 TTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGACACAAATT 907
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 TTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGACACAAATT 900

Qy     908 ATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGGATGATGG 967
           ||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db     901 ATTCAGGCTTTATGCCTAAATGTGCTAAGGTGGTGGTCTCTTCATGCACAAGGATGATGG 960

Qy     968 AGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGAACTTATA 1027
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 AGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGAACTTATA 1020

Qy    1028 TTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTATAATCTAA 1087
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 TTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTATAATCTAA 1080

Qy    1088 CAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATGTCTGGAT 1147
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 CAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATGTCTGGAT 1140

Qy    1148 TGGTTTTCCACTCACAACCAATCAATGATAGGCCAAAGCAGGCATGGTGTTGGTTTGGAG 1207
           |||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db    1141 TGGTTTTCCACTCACAACTAATCAATGATAGGCCAAAGCAGGCATGGTGTTGGTTTGGAG 1200

Qy    1208 GAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCCAGGTATA 1267
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 GAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCCAGGTATA 1260

Qy    1268 CTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGAGGAGGAGATCCGGAAG 1327
           |||||||||||||||||||||||||||||||||||| |||||||| |||||||||||||
Db    1261 CTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTAGAGGAGGAAATCCGGAAG 1320
```

```
Qy   1328 TTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATGAATTGGTTTC 1387
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 TTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATGAATTGGTTTC 1380

Qy   1388 TAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGGAACAGCATAAAAGGA 1447
          |||||| |||||||||||||||||||||||||| ||| ||||||||||||||||||||||
Db   1381 TAAATTAGGTAGAAGATAGGAATACAGCTAACCAAAAGTCAAAGGAACAGCATAAAAGGA 1440

Qy   1448 ATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAGGCAAAAATG 1507
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 ATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAGGCAAAAATG 1500

Qy   1508 TTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACCAGTCTCATAG 1567
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 TTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACCAGTCTCATAG 1560

Qy   1568 CAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTGCAGAGGTGGCAG 1627
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 CAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTGCAGAGGTGGCAG 1620

Qy   1628 AACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTCCAATTGGCTTGG 1687
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 AACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTCCAATTGGCTTGG 1680

Qy   1688 CCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCT 1747
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 CCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCT 1740

Qy   1748 TTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGT 1807
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 TTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGT 1800

Qy   1808 TGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGC 1867
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 TGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGC 1860

Qy   1868 TGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACAAAGA 1927
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 TGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACAAAGA 1920

Qy   1928 ACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATG 1987
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 ACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATG 1980

Qy   1988 CTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGCCAAATGCAAGTC 2047
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 CTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGCCAAATGCAAGTC 2040

Qy   2048 TAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGG 2107
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 TAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGG 2100

Qy   2108 AAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATG 2167
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 AAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATG 2160

Qy   2168 AATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGA 2227
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 AATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGA 2220

Qy   2228 TAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGA 2287
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 TAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGA 2280

Qy   2288 TCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTCTCTTCCC 2347
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 TCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTCTCTTCCC 2340

Qy   2348 CACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACCAGAGAAG 2407
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 CACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACCAGAGAAG 2400

Qy   2408 GCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAAT 2467
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 GCAAAGAAGGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAAT 2460

Qy   2468 ATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGCAACTGCA 2527
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 ATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGCAACTGCA 2520

Qy   2528 GAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTCTCTGCGA 2587
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 GAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTCTCTGCGA 2580

Qy   2588 CCCTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTACCTACAATATG 2641
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 CCCTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTACCTACAATATG 2634
```

```
RESULT 6
AY033233
LOCUS       AY033233               10279 bp    DNA     linear   VRL 05-JUL-2001
DEFINITION  Simian immunodeficiency virus clone 17E-Cl, complete sequence.
ACCESSION   AY033233
VERSION     AY033233.1  GI:14600346
KEYWORDS    .
SOURCE      Simian immunodeficiency virus 17E-Cl.
  ORGANISM  Simian immunodeficiency virus 17E-Cl
            Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
            lentivirus group.
REFERENCE   1  (bases 1 to 10279)
  AUTHORS   Anderson,M.G., Hauer,D., Sharma,D.P., Joag,S.V., Narayan,O.,
            Zink,M.C. and Clements,J.E.
  TITLE     Analysis of envelope changes acquired by SIVmac239 during
            neuroadaption in rhesus macaques
  JOURNAL   Virology 195 (2), 616-626 (1993)
  MEDLINE   93331720
   PUBMED   8337835
REFERENCE   2  (bases 1 to 10279)
  AUTHORS   Anderson,M.G., Hauer,D.A., Sharma,D.P., Joag,S.V., Narayan,O.,
            Zink,M.C. and Clements,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-APR-2001) Comp. Med., Johns Hopkins U. Sch. of Med.,
            600 N. Wolfe St., Baltimore, MD 21287, USA
FEATURES             Location/Qualifiers
     source          1. .10279
                     /organism="Simian immunodeficiency virus 17E-Cl"
                     /proviral
                     /db_xref="taxon:160754"
                     /clone="17E-Cl"
                     /note="recombinant of SIV/17E-Br and SIVmac239; SU portion
                     of envelope from SIV/17E-Br swarm was cloned in place of
                     the same gene in an infectious molecular clone of
                     SIVmac239"
     LTR             1. .818
                     /note="5' long terminal repeat"
     misc_feature    1. .517
                     /note="U3 region"
     misc_feature    518. .694
                     /note="R region"
     misc_feature    695. .818
                     /note="U5 region"
     gene            1053. .2585
                     /gene="gag"
     CDS             1053. .2585
                     /gene="gag"
                     /codon_start=1
                     /product="gag polyprotein"
                     /protein_id="AAK64202.1"
                     /db_xref="GI:14600350"
                     /translation="MGVRNSVLSGKKADELEKIRLRPNGKKKYMLKHVVWAANELDRF
                     GLAESLLENKEGCQKILSVLAPLVPTGSENLKSLYNTVCVIWCIHAEEKVKHTEEAKQ
                     IVQRHLVVETGTTETMPKTSRPTAPSSGRGGNYPVQQIGGNYVHLPLSPRTLNAWVKL
                     IEEKKFGAEVVPGFQALSEGCTPYDINQMLNCVGDHQAAMQIIRDIINEEAADWDLQH
                     PQPAPQQGQLREPSGSDIAGTTSSVDEQIQWMYRQQNPIPVGNIYRRWIQLGLQKCVR
```

```
                MYNPTNILDVKQGPKEPFQSYVDRFYKSLRAEQTDAAVKNWMTQTLLIQNANPDCKLV
                LKGLGVNPTLEEMLTACQGVGGPGQKARLMAEALKEALAPVPIPFAAAQQRGPRKPIK
                CWNCGKEGHSARQCRAPRRQGCWKCGKMDHVMAKCPDRQAGFLGLGPWGKKPRNFPMA
                QVHQGLMPTAPPEDPAVDLLKNYMQLGKQQREKQRESREKPYKEVTEDLLHLNSLFGG
                DQ"
gene            <2228. .5410
                /gene="pol"
CDS             <2228. .5410
                /gene="pol"
                /codon_start=1
                /product="pol polyprotein"
                /protein_id="AAK64203.1"
                /db_xref="GI:14600351"
                /translation="VLELWERGTLCKAMQSPKKTGMLEMWKNGPCYGQMPRQTGGFFR
                PWSMGKEAPQFPHGSSASGADANCSPRGPSCGSAKELHAVGQAAERKAERKQREALQG
                GDRGFAAPQFSLWRRPVVTAHIEGQPVEVLLDTGADDSIVTGIELGPHYTPKIVGGIG
                GFINTKEYKNVEIEVLGKRIKGTIMTGDTPINIFGRNLLTALGMSLNFPIAKVEPVKV
                ALKPGKDGPKLKQWPLSKEKIVALREICEKMEKDGQLEEAPPTNPYNTPTFAIKKKDK
                NKWRMLIDFRELNRVTQDFTEVQLGIPHPAGLAKRKRITVLDIGDAYFSIPLDEEFRQ
                YTAFTLPSVNNAEPGKRYIYKVLPQGWKGSPAIFQYTMRHVLEPFRKANPDVTLVQYM
                DDILIASDRTDLEHDRVVLQSKELLNSIGFSTPEEKFQKDPPFQWMGYELWPTKWKLQ
                KIELPQRETWTVNDIQKLVGVLNWAAQIYPGIKTKHLCRLIRGKMTLTEEVQWTEMAE
                AEYEENKIILSQEQEGCYYQEGKPLEATVIKSQDNQWSYKIHQEDKILKVGKFAKIKN
                THTNGVRLLAHVIQKIGKEAIVIWGQVPKFHLPVEKDVWEQWWTDYWQVTWIPEWDFI
                STPPLVRLVFNLVKDPIEGEETYYTDGSCNKQSKEGKAGYITDRGKDKVKVLEQTTNQ
                QAELEAFLMALTDSGPKANIIVDSQYVMGIITGCPTESESRLVNQIIEEMIKKSEIYV
                AWVPAHKGIGGNQEIDHLVSQGIRQVLFLEKIEPAQEEHDKYHSNVKELVFKFGLPRI
                VARQIVDTCDKCHQKGEAIHGQANSDLGTWQMDCTHLEGKIIIVAVHVASGFIEAEVI
                PQETGRQTALFLLKLAGRWPITHLHTDNGANFASQEVKMVAWWAGIEHTFGVPYNPQS
                QGVVEAMNHHLKNQIDRIREQANSVETIVLMAVHCMNFKRRGGIGDMTPAERLINMIT
                TEQEIQFQQSKNSKFKNFRVYYREGRDQLWKGPGELLWKGEGAVILKVGTDIKVVPRR
                KAKIIKDYGGGKEVDSSSHMEDTGEAREVA"
misc_feature    5084. .5099
                /gene="pol"
                /note="polypurine tract"
gene            5340. .5984
                /gene="vif"
CDS             5340. .5984
                /gene="vif"
                /codon_start=1
                /product="vif protein"
                /protein_id="AAK64204.1"
                /db_xref="GI:14600352"
                /translation="MEEEKRWIAVPTWRIPERLERWHSLIKYLKYKTKDLQKVCYVPH
                FKVGWAWWTCSRVIFPLQEGSHLEVQGYWHLTPEKGWLSTYAVRITWYSKNFWTDVTP
                NYADILLHSTYFPCFTAGEVRRAIRGEQLLSCCRFPRAHKYQVPSLQYLALKVVSDVR
                SQGENPTWKQWRRDNRRGLRMAKQNSRGDKQRGGKPPTKGANFPGLAKVLGILA"
gene            5812. .6150
                /gene="vpx"
CDS             5812. .6150
                /gene="vpx"
                /codon_start=1
                /product="vpx protein"
                /protein_id="AAK64205.1"
                /db_xref="GI:14600353"
                /translation="MSDPRERIPPGNSGEETIGEAFEWLNRTVEEINREAVNHLPREL
                IFQVWQRSWEYWHDEQGMSPSYVKYRYLCLIQKALFMHCKKGCRCLGEGHGAGGWRPG
                PPPPPPPGLA"
gene            6151. .6456
                /gene="vpr"
CDS             6151. .6456
                /gene="vpr"
                /codon_start=1
                /product="vpr protein"
                /protein_id="AAK64206.1"
                /db_xref="GI:14600354"
                /translation="MEERPPENEGPQREPWDEWVVEVLEELKEEALKHFDPRLLTALG
                NHIYNRHGDTLEGAGELIRILQRALFMHFRGGCIHSRIGQPGGGNPLSAIPPSRSML"
gene            6302. .8902
                /gene="tat"
CDS             join(6302. .6597,8803. .8902)
                /gene="tat"
                /codon_start=1
                /product="tat protein"
                /protein_id="AAK64199.1"
                /db_xref="GI:14600347"
                /translation="METPLREQENSLESSNERSSCISEADASTPESANLGEEILSQLY
                RPLEACYNTCYCKKCCYHCQFCFLKKGLGICYEQSRKRRRTPKKAKANTSSASNNRPI
                SNRTRHCQPEKAKKETVEKAVATAPGLGR"
gene            6528. .9059
                /gene="rev"
CDS             join(6528. .6597,8803. .9059)
                /gene="rev"
                /codon_start=1
                /product="rev protein"
                /protein_id="AAK64200.1"
                /db_xref="GI:14600348"
                /translation="MNNHEREEELRKRLRLIHLLHQTTDPYPTGPGTANQRRQRKRRW
                RRRWQQLLALADRIYSFPDPPTDTPLDLAIQQLQNLAIESIPDPPTNTPEALCDPTED
                SRSPQD"
gene            6604. .9243
                /gene="env"
CDS             6604. .9243
                /gene="env"
                /note="envelope glycoprotein; includes both SU and TM
                portions"
                /codon_start=1
                /product="env protein"
                /protein_id="AAK64201.1"
                /db_xref="GI:14600349"
                /translation="MGCLGNQLLIAILLLSVYGIYCTLYVTVFYGVPVWRNATIPLFC
                ATKNRDTWGTTQCLPDNGDYSEMALNVTESFDAWNNTVTEQAIEDVWQLFETSIKPCV
                KLSPLCITMRCNKSETDRWGLTKSITTTASTTSTTASARVDMVNETSSCIAQDNCAGL
                EQEQMISCKFNMTGLNRDKKKEYNETWYSADLVCEQGNNTGNESRCYMNHCNTSVIKE
                SCDKHYWDAIRFRYCAPPGYALLRCNDTNYSGFMPKCSKVVVSSCTRMMETQTSTWFG
                FNGTRAENRTYIYWHGRDNRTIISLNKYYNLTIKCRRPGNKTVLPVTIMSGLVFHSQR
                INDRPRQAWCWFGGKWKDAIKEVKQTIVKHPRYTGTNNTDKINLTAPRGGDPEVTFMW
                TNCRGEFLYCKMNWFLNWVEDRNTANQKPKEQHKRNYVPCHIRQIINTWHKVGKNVYL
                PPREGDLTCNSTVTSLIANIDWIDGNQTNITMSAEVAELYRLELGDYKLVEITPIGLA
                PTDVKRYTTGGTSRNKRGVFVLGFLGFLATAGSAMGAASLTLTAQSRTLLAGIVQQQQ
                QLLDVVKRQQELLRLTVWGTKNLQTRVTAIEKYLKDQAQLNTWGCAFRQVCHTTVPWP
                NASLIPKWNNETWQEWERKVDFLEENITALLEEAQIQQEKNMYELQKLNSWNVFGNWF
                DLASWIKYIQYGVYIVVGVILLRIVIYIVQVLAKLRQGYRPVFSSPPSYFQQTHIQQD
                PALPTREGKERDGGEGGGNSSWPWQIEYIHFLIRQLIRLLTWLFSNCRTLLSRVYQIL
                QPILQRLSATLQRIREVLRTELTYLQYGWSYFHEAVQAVWRSATETLAGAWGDLWETL
                RRGGRWILAIPRRIRQGLELTLL"
gene            9077. .9868
                /gene="nef"
                /note="contains premature stop codon"
CDS             9077. .9355
                /gene="nef"
                /codon_start=1
                /product="truncated nef protein"
                /protein_id="AAK64207.1"
                /db_xref="GI:14600355"
                /translation="MGGAISMRRSRPSGDLRQRLLRARGETYGRLLGEVEDGYSQSPG
                GLDKGLSSLSCEGQKYNQGQYMNTPWRNPAEEREKLAYRKQNMDDIDE"
misc_feature    9445. .9460
                /gene="nef"
                /note="polypurine tract"
LTR             9462. .10279
                /note="3' long terminal repeat"
misc_feature    9462. .9978
                /note="U3 region"
misc_feature    9979. .10155
                /note="R region"
polyA_signal    10132. .10137
misc_feature    10156. .10279
                /note="U5 region"
BASE COUNT      3468 a   1932 c   2569 g   2310 t
ORIGIN

  Query Match             98.9%;  Score 2612.2;  DB 14;  Length 10279;
  Best Local Similarity   99.3%;  Pred. No. 0;
  Matches 2623;  Conservative    0;  Mismatches   18;  Indels    0;  Gaps    0;
```

```
Qy      1 TCTAGAAGCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6439 TCTAGAAGCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTG 6498

Qy     61 TTTTCTTAAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCC 120
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   6499 TTTTCTTAAAAAAGGCTTGGGGATATGTTATGAACAATCACGAAAGAGAAGAAGAACTCC 6558

Qy    121 GAAAAAGGCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6559 GAAAAAGGCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGG 6618

Qy    181 AATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6619 AATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATAT 6678

Qy    241 GTCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCA 300
          |||||||||||||||||||||||| |||||||||||||||||||| |||||||||||||
Db   6679 GTCACAGTCTTTTATGGTGTACCAGTTTGGAGGAATGCGACAATTCCCCTTTTTTGTGCA 6738

Qy    301 ACCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6739 ACCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCA 6798

Qy    361 GAAGTGGCCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAG 420
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6799 GAAATGGCCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAG 6858

Qy    421 GCAATAGAGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6859 GCAATAGAGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCC 6918

Qy    481 CCATTATGCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6919 CCATTATGCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAA 6978

Qy    541 TCAATAACAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTC 600
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   6979 TCAATAACAACAACAGCATCAACAACATCAACGACAGCATCAGCAAGAGTAGACATGGTC 7038

Qy    601 AATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCAAATG 660
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   7039 AATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCGCAGGCTTGGAACAAGAGCAAATG 7098

Qy    661 ATAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAATGAA 720
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   7099 ATAAGCTGTAAATTCAACATGACAGGGTTAAATAGAGACAAGAAAAAAGAGTACAATGAA 7158

Qy    721 ACTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGA 780
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7159 ACCTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGA 7218

Qy    781 TGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGG 840
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   7219 TGTTACATGAACCACTGTAACACTTCTGTTATCAAAGAGTCTTGTGACAAACATTATTGG 7278

Qy    841 GATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGAC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7279 GATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGAC 7338

Qy    901 ACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7339 ACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGG 7398

Qy    961 ATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGA 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7399 ATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGA 7458

Qy   1021 ACTTATATTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTAT 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7459 ACTTATATTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTAT 7518

Qy   1081 AATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATG 1140
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   7519 AATCTAACAATAAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATG 7578

Qy   1141 TCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGGCCAAAGCAGGCATGGTGTTGG 1200
          |||||||||||||||||||||||| |||||||||||||||| ||||||||||||||||||
Db   7579 TCTGGATTGGTTTTCCACTCACAACGAATCAATGATAGGCCAAGGCAGGCATGGTGTTGG 7638

Qy   1201 TTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCC 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7639 TTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCC 7698

Qy   1261 AGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGAGGAGGAGAT 1320
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   7699 AGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTAGAGGAGGAGAT 7758

Qy   1321 CCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATGAAT 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7759 CCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATGAAT 7818

Qy   1381 TGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGGAACAGCAT 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7819 TGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGGAACAGCAT 7878

Qy   1441 AAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAGGC 1500
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db   7879 AAAAGGAATTACGTGCCATGTCATATCAGACAAATAATCAACACTTGGCATAAAGTAGGC 7938

Qy   1501 AAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACCAGT 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7939 AAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACCAGT 7998

Qy   1561 CTCATAGCAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTGCAGAG 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7999 CTCATAGCAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTGCAGAG 8058

Qy   1621 GTGGCAGAACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTCCAATT 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8059 GTGGCAGAACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTCCAATT 8118

Qy   1681 GGCTTGGCCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAAGA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8119 GGCTTGGCCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAAGA 8178

Qy   1741 GGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCG 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8179 GGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCG 8238

Qy   1801 GCGTCGTTGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGCAACAG 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8239 GCGTCGTTGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGCAACAG 8298

Qy   1861 CAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCTGGGGA 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8299 CAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCTGGGGA 8358

Qy   1921 ACAAAGAACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAG 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8359 ACAAAGAACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAG 8418

Qy   1981 CTGAATGCTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGCCAAAT 2040
          |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8419 CTGAATACTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGCCAAAT 8478

Qy   2041 GCAAGTCTAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGAC 2100
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   8479 GCAAGTCTAATACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGAC 8538

Qy   2101 TTCTTGGAAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAAC 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8539 TTCTTGGAAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAAC 8598

Qy   2161 ATGTATGAATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCT 2220
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
```

```
Db    8599 ATGTATGAATTACAAAAGTTGAATAGCTGGAATGTGTTTGGCAATTGGTTTGACCTTGCT 8658

Qy    2221 TCTTGGATAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGA 2280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8659 TCTTGGATAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGA 8718

Qy    2281 ATAGTGATCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTC 2340
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    8719 ATAGTGATCTATATAGTACAAGTGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTC 8778

Qy    2341 TCTTCCCCACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACC 2400
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8779 TCTTCCCCACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACC 8838

Qy    2401 AGAGAAGGCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAG 2460
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8839 AGAGAAGGCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAG 8898

Qy    2461 ATAGAATATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGC 2520
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8899 ATAGAATATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGC 8958

Qy    2521 AACTGCAGAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTC 2580
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8959 AACTGCAGAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTC 9018

Qy    2581 TCTGCGACCCTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTACCTACAATAT 2640
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9019 TCTGCGACCCTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTACCTACAATAT 9078

Qy    2641 G 2641
           |
Db    9079 G 9079
```

```
RESULT 7
SIVMM316ZY
LOCUS       SIVMM316ZY              2799 bp ss-DNA     linear   VRL 21-DEC-1993
DEFINITION  Simian immunodeficiency virus (Mm316zl) envelope glycoprotein and
            nef protein genes, complete cds and 5'end.
ACCESSION   L22822
VERSION     L22822.1  GI:431596
KEYWORDS    envelope glycoprotein; nef protein.
SOURCE      Simian immunodeficiency virus DNA.
  ORGANISM  Simian immunodeficiency virus
            Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
            lentivirus group.
REFERENCE   1  (bases 1 to 2799)
  AUTHORS   Kodama,T., Mori,K., Kawahara,T., Ringler,D.J. and Desrosiers,R.C.
  TITLE     Analysis of simian immunodeficiency virus sequence variation in
            tissues of rhesus macaques with simian AIDS
  JOURNAL   J. Virol. 67 (11), 6522-6534 (1993)
  MEDLINE   94016835
   PUBMED   8411355
FEATURES             Location/Qualifiers
     source          1. .2799
                     /organism="Simian immunodeficiency virus"
                     /proviral
                     /db_xref="taxon:11723"
     gene            1. .2457
                     /gene="tat"
     exon            1. .150
                     /gene="tat"
                     /number=1
     gene            83. .2614
                     /gene="rev"
     exon            83. .150
                     /gene="tat"
                     /number=1
     misc_signal     150
                     /gene="tat"
                     /note="splice donor site"
     gene            159. .2798
                     /gene="env"
     CDS             159. .2798
                     /gene="env"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAA47706.1"
                     /db_xref="GI:431597"
                     /translation="MGCLGNQLLIAILLLSVYGIYCTLYVTVFYGVPAWRNATIPLFC
                     ATKNRDTWGTTQCLPDNGDYSEMALNVTESFDAWNNTVIEQAIEDVWQLFETSIKPCV
                     KLSPLCITMRCNKSETDRWGLTKSITTTASTTSTTASAKVDMVNETSSCIAQDNCTGL
                     EQEQMISCKFNMTGLKRDKKKEYNETWYSADLVCEQGNNTGNESRCYMNHCNTSVIQE
                     SCDKHYWDAIRFRYCAPPGYALLRCNDTNYSGFMPKCSKVVVSSCTRMMETQTSTWFG
                     FNGTRAENRTYIYWHGRDNRTIISLNKYYNLTMKCRRPGNKTVLPVTIMSGLVFHSQR
                     INDRPKQAWCWFGGKWKDAIKEVKQTIVKHPRYTGTNNTDKINLTAPRGGNPEATFMW
                     TNCRGEFLYCKMNWFLNWVEDRNTANQKSKEQHKRNYVPCHIRQIINTWHKVGKNVYL
                     PPREGDLTCNSTVTSLIANIDWIDGNQTNITMSAEVAELYRLELGDYKLVEITPIGLA
                     PTDVKRYTTGGTSRNKRGVFVLGFLGFLATAGSAMGAASLTLTAQSRTLLAGIVQQQQ
                     QLLDVVKRQQELLRLTVWGTKNLQTRVTAIEKYLKDQAQLNAWGCAFRQVCHTTVPWP
                     NASLTPKWNNETWQEWERKVDFLEENITALLEEAQIQQEKNTYELQKLNSWDVFGNWF
                     DLASWIKYIQYGVYIVVGVILLRIVIYIVQMLAKLRQGYRPVFSSPPSYFQQTHIQQD
                     PALPTREGKEGDGGEGGGNSSWPWQIEYIHFLIRQLIRLLTWLFSNCRTLLSRVYQIL
                     QPILQRLSATLQRIREVLRTELTYLQYGWSYFHEAVQAVWRSATETLAGAWGDLWETL
                     RRGGRWILAIPRRIRQGLELTLL"
     misc_signal     1734
                     /gene="tat"
                     /note="SU/TM cleavage site"
     exon            2358. .2614
                     /gene="rev"
                     /number=2
     exon            2358. .2457
                     /gene="tat"
                     /number=2
     misc_signal     2358
                     /gene="tat"
                     /note="splice acceptor"
     gene            2632. .2799
                     /gene="nef"
     CDS             2632. .>2799
                     /gene="nef"
                     /codon_start=1
                     /product="nef protein"
                     /protein_id="AAA47707.1"
                     /db_xref="GI:431598"
                     /translation="MGGAISMRRSRPSGDLRQRLLRARGETYGRLLGEVEDGYSQSPG
                     GLDKGLSSLSCE"
BASE COUNT      899 a    525 c    673 g    702 t
ORIGIN
```

```
  Query Match             98.9%;  Score 2611.6;  DB 14;  Length 2799;
  Best Local Similarity   99.5%;  Pred. No. 0;
  Matches 2620;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy       8 GCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTGTTTTCTT 67
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTGTTTTCTT 60

Qy      68 AAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCCGAAAAAG 127
           |||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db      61 AAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCCGGAAAAA 120

Qy     128 GCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGGAATCAGC 187
           |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db     121 GCTAAGGCTAATACATCTTCTGCATCAAGCAAGTAAGTATGGGATGTCTTGGGAATCAGC 180

Qy     188 TGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATATGTCACAG 247
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATATGTCACAG 240

Qy     248 TCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCAACCAAGA 307
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCAACCAAGA 300

Qy     308 ATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCAGAAGTGG 367
```

```
Db    301 ATAGGGATACTTGGGGGACAACTCAGTGCCTACCAGATAATGGTGATTATTCAGAAATGG 360

Qy    368 CCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAGGCAATAG 427
          ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db    361 CCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCATAGAACAGGCAATAG 420

Qy    428 AGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCCCCCATTAT 487
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCCCCCATTAT 480

Qy    488 GCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAATCAATAA 547
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAATCAATAA 540

Qy    548 CAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTCAATGAGA 607
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTCAATGAGA 600

Qy    608 CTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCAAATGATAAGCT 667
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCAAATGATAAGCT 660

Qy    668 GTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAATGAAACTTGGT 727
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAATGAAACTTGGT 720

Qy    728 ACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGATGTTACA 787
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 ACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGATGTTACA 780

Qy    788 TGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGGGATGCTA 847
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 TGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGGGATGCTA 840

Qy    848 TTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGACACAAATT 907
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 TTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGACACAAATT 900

Qy    908 ATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGGATGATGG 967
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 ATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGGATGATGG 960

Qy    968 AGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGAACTTATA 1027
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 AGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGAACTTATA 1020

Qy   1028 TTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTATAATCTAA 1087
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 TTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTATAATCTAA 1080

Qy   1088 CAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATGTCTGGAT 1147
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATGTCTGGAT 1140

Qy   1148 TGGTTTTCCACTCACAACCAATCAATGATAGGCCAAAGCAGGCATGGTGTTGGTTTGGAG 1207
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   1141 TGGTTTTCCACTCACAACGAATCAATGATAGGCCAAAGCAGGCATGGTGTTGGTTTGGAG 1200

Qy   1208 GAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCCAGGTATA 1267
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 GAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCCAGGTATA 1260

Qy   1268 CTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGAGGAGGAGATCCGGAAG 1327
          |||||||||||||||||||||||||||||||||||||||| |||||||| ||||||||||
Db   1261 CTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTAGAGGAGGAAATCCGGAAG 1320

Qy   1328 TTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATGAATTGGTTTC 1387
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 CTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATGAATTGGTTTC 1380

Qy   1388 TAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGGAACAGCATAAAAGGA 1447
          ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
```

```
Db   1381 TAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGTCAAAGGAACAGCATAAAAGGA 1440

Qy   1448 ATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAGGCAAAAATG 1507
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 ATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAGGCAAAAATG 1500

Qy   1508 TTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACCAGTCTCATAG 1567
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 TTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACCAGTCTCATAG 1560

Qy   1568 CAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTGCAGAGGTGGCAG 1627
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 CAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTGCAGAGGTGGCAG 1620

Qy   1628 AACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTCCAATTGGCTTGG 1687
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 AACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTCCAATTGGCTTGG 1680

Qy   1688 CCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCT 1747
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 CCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCT 1740

Qy   1748 TTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGT 1807
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 TTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGT 1800

Qy   1808 TGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGC 1867
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 TGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGC 1860

Qy   1868 TGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACAAAGA 1927
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 TGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACAAAGA 1920

Qy   1928 ACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATG 1987
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 ACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATG 1980

Qy   1988 CTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGCCAAATGCAAGTC 2047
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 CTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGCCAAATGCAAGTC 2040

Qy   2048 TAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGG 2107
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 TAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGG 2100

Qy   2108 AAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATG 2167
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   2101 AGGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACACGTATG 2160

Qy   2168 AATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGA 2227
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 AATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGA 2220

Qy   2228 TAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGA 2287
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 TAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGA 2280

Qy   2288 TCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTCTCTTCCC 2347
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 TCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTCTCTTCCC 2340

Qy   2348 CACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACCAGAGAAG 2407
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 CACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACCAGAGAAG 2400

Qy   2408 GCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAAT 2467
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 GCAAAGAAGGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAAT 2460

Qy   2468 ATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGCAACTGCA 2527
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 ATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGCAACTGCA 2520
```

```
Qy    2528 GAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTCTCTGCGA 2587
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2521 GAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTCTCTGCGA 2580

Qy    2588 CCCTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTACCTACAATATG 2641
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2581 CCCTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTACCTACAATATG 2634


RESULT 8
SIVMM316ZT
LOCUS       SIVMM316ZT              2799 bp ss-DNA     linear   VRL 21-DEC-1993
DEFINITION  Simian immunodeficiency virus (Mm316zh) envelope glycoprotein and
            nef protein genes, complete cds and 5' end.
ACCESSION   L22817
VERSION     L22817.1  GI:431582
KEYWORDS    envelope glycoprotein; nef protein.
SOURCE      Simian immunodeficiency virus DNA.
  ORGANISM  Simian immunodeficiency virus
            Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
            lentivirus group.
REFERENCE   1  (bases 1 to 2799)
  AUTHORS   Kodama,T., Mori,K., Kawahara,T., Ringler,D.J. and Desrosiers,R.C.
  TITLE     Analysis of simian immunodeficiency virus sequence variation in
            tissues of rhesus macaques with simian AIDS
  JOURNAL   J. Virol. 67 (11), 6522-6534 (1993)
  MEDLINE   94016835
   PUBMED   8411355
FEATURES             Location/Qualifiers
     source          1. .2799
                     /organism="Simian immunodeficiency virus"
                     /proviral
                     /db_xref="taxon:11723"
     gene            1. .2457
                     /gene="tat"
     exon            1. .150
                     /gene="tat"
                     /number=1
     gene            83. .2614
                     /gene="rev"
     exon            83. .150
                     /gene="tat"
                     /number=1
     misc_signal     150
                     /gene="tat"
                     /note="splice donor site"
     gene            159. .2798
                     /gene="env"
     CDS             159. .2798
                     /gene="env"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAA47698.1"
                     /db_xref="GI:431583"
                     /translation="MGCLGNQLLIAILLLSVYGIYCTLYVTVFYGVPVWRNATIPLFC
                     ATKNRDTWGTTQCLPDNGDYSEMALNVTESFDAWNNTVTEQAIEDVWQLFETSIKPCV
                     KLSPLCITMRCNKSETDRWGLTKSITTTASTTSTTASAKVDMVNETSSCIAQDNCAGL
                     EQEQMISCKFNMTGLKRDKKKEYNETWYSADLVCEQGNNTGNESRCYMNHCNTSVIHE
                     SCDKHYWDAIRFRYCAPPGYALLRCNDTNYSGFMPKCSKVVVSSCTRMMETQTSTWFG
                     FNGTSAENRTYIYWHGRDNRTIISLNKYYNLTMKCRRPGNKTVLPVTIMSGLVFHSQR
                     INDRPKQAWCWFGGKWKDAIKEVKQTIVKHPRYTGTNNTDKINLTAPRGGNPEATFMW
                     TNCRGESLYCKMNWFLNWVEDRNTANQKSKEQHKRNYVPCHIRQIINTWHKVGKNVYL
                     PPREGDLTCNSTVTSLIANIDWIDGNQTNITMSAEVAELYRLELGDYKLVEITPIGLA
                     PTDVKRYTTGGTSRNKRGVFVLGFLGFLATAGSAMGAASLTLTAQSRTLLAGIVQQQQ
                     QLLDVVKRQQELLRLTVWGTKNLQTRVTAIEKYLKDQAQLNAWGCAFRQVCHTTVPWP
                     NASLTPKWNNETWQEWERKVDFLEENITALLEEAQIQQEKNMYELQKLNSWDVFGNWF
                     DLASWIKYIQYGVYIVVGVILLRIVIYIVQMLAKLRQGYRPVFSSPPSYFQQTHIQQD
                     PALPTREGKEGDGGEGGNSSWPWQIEYIHFLIRQLIRLLTWLFSNCRTLLSRVYQIL
                     QPILQRLSATLQRIREVLRTELTYLQYGWSYFHEAVQAVWRSATETLAGAWGDLWETL
                     RRGGRWILAIPRRIRQGLELTLL"
     misc_signal     1734
                     /gene="tat"
                     /note="SU/TM cleavage site"
     exon            2358. .2614
                     /gene="rev"
                     /number=2
     exon            2358. .2457
                     /gene="tat"
                     /number=2
     misc_signal     2358
                     /gene="tat"
                     /note="splice acceptor"
     gene            2632. .2799
                     /gene="nef"
     CDS             2632. .>2799
                     /gene="nef"
                     /codon_start=1
                     /product="nef protein"
                     /protein_id="AAA47699.1"
                     /db_xref="GI:431584"
                     /translation="MGGAISMRRSRPSGDLRQRLLRARGETYGRLLGEVEDGYSQSPG
                     GLDKGLSSLSCE"
BASE COUNT      899 a    524 c    671 g    705 t
ORIGIN

  Query Match             98.8%;  Score 2610;  DB 14;  Length 2799;
  Best Local Similarity   99.4%;  Pred. No. 0;
  Matches 2619;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy      8 GCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTGTTTTCTT 67
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTGTTTTCTT 60

Qy     68 AAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCCGAAAAAG 127
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCCGAAAAAG 120

Qy    128 GCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGGAATCAGC 187
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGGAATCAGC 180

Qy    188 TGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATATGTCACAG 247
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATATGTCACAG 240

Qy    248 TCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCAACCAAGA 307
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    241 TCTTTTATGGTGTACCAGTTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCAACCAAGA 300

Qy    308 ATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCAGAAGTGG 367
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    301 ATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCAGAAATGG 360

Qy    368 CCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAGGCAATAG 427
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAGGCAATAG 420

Qy    428 AGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCCCCATTAT 487
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCCCCATTAT 480

Qy    488 GCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAATCAATAA 547
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAATCAATAA 540

Qy    548 CAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTCAATGAGA 607
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTCAATGAGA 600

Qy    608 CTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCAAATGATAAGCT 667
          |||||||||||||||||| ||||||||| ||||||||||||||||||||||||||||||
Db    601 CTAGTTCTTGTATAGCCCAAGATAATTGCGCAGGCTTGGAACAAGAGCAAATGATAAGCT 660

Qy    668 GTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAATGAAACTTGGT 727
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    661 GTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAATGAAACTTGGT 720

Qy    728 ACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGATGTTACA 787
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 ACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGATGTTACA 780

Qy    788 TGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGGGATGCTA 847
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db    781 TGAACCACTGTAACACTTCTGTTATCCATGAGTCTTGTGACAAACATTATTGGGATGCTA 840

Qy    848 TTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGACACAAATT 907
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 TTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGACACAAATT 900

Qy    908 ATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGGATGATGG 967
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 ATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGGATGATGG 960

Qy    968 AGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGAACTTATA 1027
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    961 AGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGTGCAGAAAATAGAACTTATA 1020

Qy   1028 TTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTATAATCTAA 1087
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 TTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTATAATCTAA 1080

Qy   1088 CAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATGTCTGGAT 1147
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATGTCTGGAT 1140

Qy   1148 TGGTTTTCCACTCACAACCAATCAATGATAGGCCAAAGCAGGCATGGTGTTGGTTTGGAG 1207
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   1141 TGGTTTTCCACTCACAACGAATCAATGATAGGCCAAAGCAGGCATGGTGTTGGTTTGGAG 1200

Qy   1208 GAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCCAGGTATA 1267
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 GAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCCAGGTATA 1260

Qy   1268 CTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGAGGAGGAGATCCGGAAG 1327
          |||||||||||||||||||||||||||||||||| |||||||| ||||||||
Db   1261 CTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTAGAGGAGGAAATCCGGAAG 1320

Qy   1328 TTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATGAATTGGTTTC 1387
           |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db   1321 CTACCTTCATGTGGACAAATTGCAGAGGAGAGTCCCTCTACTGTAAAATGAATTGGTTTC 1380

Qy   1388 TAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGGAACAGCATAAAAGGA 1447
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   1381 TAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGTCAAAGGAACAGCATAAAAGGA 1440

Qy   1448 ATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAGGCAAAAATG 1507
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 ATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAGGCAAAAATG 1500

Qy   1508 TTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACCAGTCTCATAG 1567
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 TTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACCAGTCTCATAG 1560

Qy   1568 CAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTGCAGAGGTGGCAG 1627
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 CAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTGCAGAGGTGGCAG 1620

Qy   1628 AACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTCCAATTGGCTTGG 1687
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 AACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTCCAATTGGCTTGG 1680

Qy   1688 CCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCT 1747
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 CCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCT 1740

Qy   1748 TTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGT 1807
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 TTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGT 1800
```

```
Qy   1808 TGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGC 1867
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   1801 TGACGCTGACCGCTCAGTCCCGGACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGC 1860

Qy   1868 TGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACAAAGA 1927
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 TGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACAAAGA 1920

Qy   1928 ACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATG 1987
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 ACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATG 1980

Qy   1988 CTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGCCAAATGCAAGTC 2047
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 CTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGCCAAATGCAAGTC 2040

Qy   2048 TAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGG 2107
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 TAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGG 2100

Qy   2108 AAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATG 2167
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 AAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATG 2160

Qy   2168 AATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGA 2227
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 AATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGA 2220

Qy   2228 TAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGA 2287
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 TAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGA 2280

Qy   2288 TCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTCTCTTCCC 2347
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 TCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTCTCTTCCC 2340

Qy   2348 CACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACCAGAGAAG 2407
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 CACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACCAGAGAAG 2400

Qy   2408 GCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAAT 2467
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 GCAAAGAAGGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAAT 2460

Qy   2468 ATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGCAACTGCA 2527
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 ATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGCAACTGCA 2520

Qy   2528 GAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTCTCTGCGA 2587
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 GAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTCTCTGCGA 2580

Qy   2588 CCCTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTACCTACAATATG 2641
          || ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 CCTTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTACCTACAATATG 2634
```

```
RESULT 9
SIVMM316ZV
LOCUS       SIVMM316ZV              2799 bp ss-DNA     linear   VRL 21-DEC-1993
DEFINITION  Simian immunodeficiency virus (Mm316zj) envelope glycoprotein and
            nef protein genes, complete cds and 5'end.
ACCESSION   L22819
VERSION     L22819.1  GI:431587
KEYWORDS    envelope glycoprotein; nef protein.
SOURCE      Simian immunodeficiency virus DNA.
  ORGANISM  Simian immunodeficiency virus
            Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
            lentivirus group.
REFERENCE   1  (bases 1 to 2799)
  AUTHORS   Kodama,T., Mori,K., Kawahara,T., Ringler,D.J. and Desrosiers,R.C.
  TITLE     Analysis of simian immunodeficiency virus sequence variation in
```

```
          tissues of rhesus macaques with simian AIDS
  JOURNAL   J. Virol. 67 (11), 6522-6534 (1993)
  MEDLINE   94016835
   PUBMED   8411355
FEATURES             Location/Qualifiers
     source          1. .2799
                     /organism="Simian immunodeficiency virus"
                     /proviral
                     /db_xref="taxon:11723"
     gene            1. .2457
                     /gene="tat"
     exon            1. .150
                     /gene="tat"
                     /number=1
     gene            83. .2614
                     /gene="rev"
     exon            83. .150
                     /gene="tat"
                     /number=1
     misc_signal     150
                     /gene="tat"
                     /note="splice donor site"
     gene            159. .2798
                     /gene="env"
     CDS             159. .2798
                     /gene="env"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAA47700.1"
                     /db_xref="GI:431588"
                     /translation="MGCLGNQLLIAILLLSVYRIYCTLYVTVFYGVPVWRNATIPLFC
                     ATKNRDTWGTTQCLPDNGDYSEMALNVTESFDAWNNTVTEQAIEDVWQLFETSIKPCA
                     KLSPLCITMRCNKSETDRWGLTKSITTTASTTSTTASAKVDMVNEASSCIAQDNCAGL
                     EQEQMISCKFNMTGLKRNKKKEYNETWYSADLVCEQGNNTGNESRCYMNHCNTSVIQE
                     SCDKHYWDAIRFRYCAPPGYALLRCNDTNYSGFMPKCSKVVASSCTRMMETQTSTWFG
                     FNGTRAENRTYIYWHGRDNRTIISLNKYYNLTMKCRRPGNKTVLPVTIMSGLVFHSQL
                     INDRPKQAWCWFGGKWKDAIKEVKQTIVKHPRYTGTNNTDKINLTAPRGGNPEATFMW
                     TNCRGEFLYCKMNWFLNWVEDRNTANQKSKEQHKRNYVPCHIRQIINTWHKVGKNVYL
                     PPREGDLTCNSTVTSLIANIDWIDGNQTNITMSAEVAELYRLELGDYKLVEITPIGLA
                     PTDVKRYTTGGTSRNKRGVFVLGFLGFLATAGSAMGAASLTLTAQSRTLLAGIVQQQQ
                     QLLDVVKRQQELLRLTVWGTKNLQTRVTAIEKYLKDQAQLNAWGCAFRQVCHTTVPWP
                     NASLTPKWNNETWQEWERKVDFLEENITALLEEAQIQQEKNMYELQKLNSWDVFGNWF
                     DLASWIKYIQYGVYIVVGVILLRIVIYIVQMLAKLRQGYRPVFSSPPSYFQQTHIQQD
                     PALPTREGKEGDGGEGGGNSSWPWQIEYIHFLIRQLIRLLTWLFSNCRTLLSRVYQIL
                     QPILQRLSATLQRIREVLRTELTYLQYGWSYFHEAVQAVWRSATETLAGAWGDLWETL
                     RRGGRWILAIPRRIRQGLELTLL"
     misc_signal     1734
                     /gene="tat"
                     /note="SU/TM cleavage site"
     exon            2358. .2614
                     /gene="rev"
                     /number=2
     exon            2358. .2457
                     /gene="tat"
                     /number=2
     misc_signal     2358
                     /gene="tat"
                     /note="splice acceptor"
     gene            2632. .2799
                     /gene="nef"
     CDS             2632. .>2799
                     /gene="nef"
                     /codon_start=1
                     /product="nef protein"
                     /protein_id="AAA47701.1"
                     /db_xref="GI:431589"
                     /translation="MGGAISMRRSRPSGDLRQRLLRARGETYGRLLGEVEDGYSQSPG
                     GLDKGLSSLSCE"
BASE COUNT      903 a    525 c    669 g    702 t
ORIGIN

  Query Match             98.8%;  Score 2610;  DB 14;  Length 2799;
  Best Local Similarity   99.4%;  Pred. No. 0;

  Matches 2619;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy       8 GCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTGTTTTCTT 67
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTGTTTTCTT 60

Qy      68 AAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCCGAAAAAG 127
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCCGAAAAAG 120

Qy     128 GCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGGAATCAGC 187
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGGAATCAGC 180

Qy     188 TGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATATGTCACAG 247
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     181 TGCTTATCGCCATCTTGCTTTTAAGTGTCTATAGGATCTATTGTACTCTATATGTCACAG 240

Qy     248 TCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCAACCAAGA 307
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     241 TCTTTTATGGTGTACCAGTTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCAACCAAGA 300

Qy     308 ATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCAGAAGTGG 367
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db     301 ATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCAGAAATGG 360

Qy     368 CCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAGGCAATAG 427
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAGGCAATAG 420

Qy     428 AGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCCCCATTAT 487
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     421 AGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGCAAAATTATCCCCATTAT 480

Qy     488 GCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAATCAATAA 547
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 GCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAATCAATAA 540

Qy     548 CAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTCAATGAGA 607
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 CAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTCAATGAGG 600

Qy     608 CTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCAAATGATAAGCT 667
          |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db     601 CTAGTTCTTGTATAGCCCAGGATAATTGCGCAGGCTTGGAACAAGAGCAAATGATAAGCT 660

Qy     668 GTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAATGAAACTTGGT 727
          |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db     661 GTAAATTCAACATGACAGGGTTAAAAAGAAACAAGAAAAAAGAGTACAATGAAACTTGGT 720

Qy     728 ACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGATGTTACA 787
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 ACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGATGTTACA 780

Qy     788 TGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGGGATGCTA 847
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 TGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGGGATGCTA 840

Qy     848 TTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGACACAAATT 907
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 TTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGACACAAATT 900

Qy     908 ATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGGATGATGG 967
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     901 ATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGCCTCTTCATGCACAAGGATGATGG 960

Qy     968 AGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGAACTTATA 1027
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db     961 AGACACAGACTTCTACTTGGTTTGGATTTAATGGAACTAGAGCAGAAAATAGAACTTATA 1020

Qy    1028 TTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTATAATCTAA 1087
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 TTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTATAATCTAA 1080
```

```
Qy   1088 CAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATGTCTGGAT 1147
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATGTCTGGAT 1140

Qy   1148 TGGTTTTCCACTCACAACCAATCAATGATAGGCCAAAGCAGGCATGGTGTTGGTTTGGAG 1207
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   1141 TGGTTTTCCACTCACAACTAATCAATGATAGGCCAAAGCAGGCATGGTGTTGGTTTGGAG 1200

Qy   1208 GAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCCAGGTATA 1267
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 GAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCCAGGTATA 1260

Qy   1268 CTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGAGGAGGAGATCCGGAAG 1327
          |||||||||||||||||||||||||||||||||| |||||||| |||||||||
Db   1261 CTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTAGAGGAGGAAATCCGGAAG 1320

Qy   1328 TTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATGAATTGGTTTC 1387
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 CTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATGAATTGGTTTC 1380

Qy   1388 TAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGGAACAGCATAAAAGGA 1447
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   1381 TAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGTCAAAGGAACAGCATAAAAGGA 1440

Qy   1448 ATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAGGCAAAAATG 1507
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 ATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAGGCAAAAATG 1500

Qy   1508 TTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACCAGTCTCATAG 1567
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 TTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACCAGTCTCATAG 1560

Qy   1568 CAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTGCAGAGGTGGCAG 1627
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 CAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTGCAGAGGTGGCAG 1620

Qy   1628 AACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTCCAATTGGCTTGG 1687
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 AACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTCCAATTGGCTTGG 1680

Qy   1688 CCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCT 1747
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 CCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCT 1740

Qy   1748 TTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGT 1807
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 TTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGT 1800

Qy   1808 TGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGC 1867
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 TGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGC 1860

Qy   1868 TGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACAAAGA 1927
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 TGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACAAAGA 1920

Qy   1928 ACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATG 1987
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 ACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATG 1980

Qy   1988 CTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGCCAAATGCAAGTC 2047
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 CTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGCCAAATGCAAGTC 2040

Qy   2048 TAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGG 2107
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 TAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGG 2100

Qy   2108 AAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATG 2167
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 AAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATG 2160

Qy   2168 AATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGA 2227
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 AATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGA 2220

Qy   2228 TAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGA 2287
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 TAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGA 2280

Qy   2288 TCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTCTCTTCCC 2347
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 TCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTCTCTTCCC 2340

Qy   2348 CACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACCAGAGAAG 2407
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 CACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACCAGAGAAG 2400

Qy   2408 GCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAAT 2467
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 GCAAAGAAGGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAAT 2460

Qy   2468 ATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGCAACTGCA 2527
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 ATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGCAACTGCA 2520

Qy   2528 GAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTCTCTGCGA 2587
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 GAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTCTCTGCGA 2580

Qy   2588 CCCTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTACCTACAATATG 2641
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 CCCTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTACCTACAATATG 2634
```

```
RESULT 10
SIVMM316ZX
LOCUS       SIVMM316ZX              2799 bp ss-DNA     linear   VRL 21-DEC-1993
DEFINITION  Simian immunodeficiency virus (Mm132zk) envelope glycoprotein and
            nef protein genes, complete cds and 5'end.
ACCESSION   L22821
VERSION     L22821.1  GI:431593
KEYWORDS    envelope glycoprotein; nef protein.
SOURCE      Simian immunodeficiency virus DNA.
  ORGANISM  Simian immunodeficiency virus
            Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
            lentivirus group.
REFERENCE   1  (bases 1 to 2799)
  AUTHORS   Kodama,T., Mori,K., Kawahara,T., Ringler,D.J. and Desrosiers,R.C.
  TITLE     Analysis of simian immunodeficiency virus sequence variation in
            tissues of rhesus macaques with simian AIDS
  JOURNAL   J. Virol. 67 (11), 6522-6534 (1993)
  MEDLINE   94016835
   PUBMED   8411355
FEATURES             Location/Qualifiers
     source          1. .2799
                     /organism="Simian immunodeficiency virus"
                     /proviral
                     /db_xref="taxon:11723"
     gene            1. .2457
                     /gene="tat"
     exon            1. .150
                     /gene="tat"
                     /number=1
     gene            83. .2614
                     /gene="rev"
     exon            83. .150
                     /gene="tat"
                     /number=1
     misc_signal     150
                     /gene="tat"
                     /note="splice donor site"
     gene            159. .2798
                     /gene="env"
     CDS             159. .2798
```

```
                     /gene="env"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAA47704.1"
                     /db_xref="GI:431594"
                     /translation="MGCLGNQLLIAILLLSVYGIYCTLYVTVFYGVPAWRNATIPLFC
                     ATKNRDTWGTTQCLPDNGDYSEMALNVTESFDAWNNTVTEQAIEDVWQLLETSIKPCV
                     KLSPLCITMRCNKSETDRWGLTKSITTTASTTSTTASAKVDMVNETSSCIAQDNCAGL
                     EQEQMISCKFNMTGLERNKKKEYNETWYSADLVCEQGNNTGNESRCYMNHCNTSVIQE
                     SCDKHYWDAIRFRYCAPPGYALLRCNDTNYSGFMPKCSKVVVSSCTRMMETQTSTWFG
                     FNGTRAENRTYIYWHGRNNRTIISLNKYYNLTIKCRRPGNKTVLPVTIMSGLVFHSQP
                     INDRPKQAWCWFGGKWKDAIKEVKQTIVKHPRYTGTNNTDKINLTAPRGGNPEATFMW
                     TNCRGEFLYCKMNWFLNWVEDRNTANQKSKEQHKRNYVPCHIRQIINTWHKVGKNVYL
                     PPREGDLTCNSTVTSLIANIDWIDGNQTNITMSAEVAELYRLELGDYKLVEITPIGLA
                     PTDVKRYTTGGTSRNKRGVFVLGFLGFLATAGSAMGAASLTLTAQSRTLLAGIVQQQQ
                     QLLDVVTRQQELLRLTVWGTKNLQTRVTAIEKYLKDQAQLNAWGCAFRQVCHTTVPWP
                     NASLTPKWNNETWQEWERKVDFLEENITALLEEAQIQQEKNMYELQKLNSWDVFGNWF
                     DLASWIKYIQYGVYIVVGVILLRIVIYIVQMLAKLRQGYRPVFSSPPSYFQQTHIQQD
                     PALPTREGKEGDGGEGGGNSSWPWQIEYIHFLIRQLIRLLTWLFSNCRTLLSRVYQIL
                     QPILQRLSATLQRIREVLRTELTYLQYGWSYFHEAVQAVWRSATETLAGAWGDLWETL
                     RRGGRWILAIPRRIRQGLELTLL"
     misc_signal     1734
                     /gene="tat"
                     /note="SU/TM cleavage site"
     exon            2358. .2614
                     /gene="rev"
                     /number=2
     exon            2358. .2457
                     /gene="tat"
                     /number=2
     misc_signal     2358
                     /gene="tat"
                     /note="splice acceptor"
     gene            2632. .2799
                     /gene="nef"
     CDS             2632. .>2799
                     /gene="nef"
                     /codon_start=1
                     /product="nef protein"
                     /protein_id="AAA47705.1"
                     /db_xref="GI:431595"
                     /translation="MGGAISMRRSRPSGDLRQRLLRARGETYGRLLGEVEDGYSQSPG
                     GLDKGLSSLSCE"
BASE COUNT      903 a    529 c    667 g    700 t
ORIGIN

  Query Match             98.8%;  Score 2610;  DB 14;  Length 2799;
  Best Local Similarity   99.4%;  Pred. No. 0;
  Matches 2619;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy      8 GCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTGTTTTCTT 67
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTGTTTTCTT 60

Qy     68 AAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCCGAAAAAG 127
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCCGAAAAAG 120

Qy    128 GCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGGAATCAGC 187
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGGAATCAGC 180

Qy    188 TGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATATGTCACAG 247
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATATGTCACAG 240

Qy    248 TCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCAACCAAGA 307
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCAACCAAGA 300

Qy    308 ATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCAGAAGTGG 367
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    301 ATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCAGAAATGG 360

Qy    368 CCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAGGCAATAG 427
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAGGCAATAG 420

Qy    428 AGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCCCCATTAT 487
          ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    421 AGGATGTATGGCAACTCCTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCCCCATTAT 480

Qy    488 GCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAATCAATAA 547
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAATCAATAA 540

Qy    548 CAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTCAATGAGA 607
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTCAATGAGA 600

Qy    608 CTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCAAATGATAAGCT 667
          ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db    601 CTAGTTCTTGTATAGCCCAGGATAATTGCGCAGGCTTGGAACAAGAGCAAATGATAAGCT 660

Qy    668 GTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAATGAAACTTGGT 727
          ||||||||||||||||||||||| ||||| ||||||||||||||||||||||||||||||
Db    661 GTAAATTCAACATGACAGGGTTAGAAAGAAACAAGAAAAAAGAGTACAATGAAACTTGGT 720

Qy    728 ACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGATGTTACA 787
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 ACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGATGTTACA 780

Qy    788 TGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGGGATGCTA 847
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 TGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGGGATGCTA 840

Qy    848 TTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGACACAAATT 907
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 TTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGACACAAATT 900

Qy    908 ATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGGATGATGG 967
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 ATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGGATGATGG 960

Qy    968 AGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGAACTTATA 1027
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 AGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGAACTTATA 1020

Qy   1028 TTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTATAATCTAA 1087
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   1021 TTTACTGGCATGGTAGGAATAATAGGACTATAATTAGTTTAAATAAGTATTATAATCTAA 1080

Qy   1088 CAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATGTCTGGAT 1147
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CAATAAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATGTCTGGAT 1140

Qy   1148 TGGTTTTCCACTCACAACCAATCAATGATAGGCCAAAGCAGGCATGGTGTTGGTTTGGAG 1207
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 TGGTTTTCCACTCACAACCAATCAATGATAGGCCAAAGCAGGCATGGTGTTGGTTTGGAG 1200

Qy   1208 GAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCCAGGTATA 1267
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 GAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCCAGGTATA 1260

Qy   1268 CTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGAGGAGGAGATCCGGAAG 1327
          |||||||||||||||||||||||||||||||||||||||| |||||||| | |||||||
Db   1261 CTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTAGAGGAGGAAACCCGGAAG 1320

Qy   1328 TTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATGAATTGGTTTC 1387
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 CTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATGAATTGGTTTC 1380

Qy   1388 TAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGGAACAGCATAAAAGGA 1447
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   1381 TAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGTCAAAGGAACAGCATAAAAGGA 1440
```

```
Qy   1448 ATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAGGCAAAAATG 1507
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 ATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAGGCAAAAATG 1500

Qy   1508 TTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACCAGTCTCATAG 1567
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 TTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACCAGTCTCATAG 1560

Qy   1568 CAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTGCAGAGGTGGCAG 1627
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 CAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTGCAGAGGTGGCAG 1620

Qy   1628 AACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTCCAATTGGCTTGG 1687
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 AACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTCCAATTGGCTTGG 1680

Qy   1688 CCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCT 1747
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 CCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCT 1740

Qy   1748 TTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGT 1807
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 TTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGT 1800

Qy   1808 TGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGC 1867
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 TGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGC 1860

Qy   1868 TGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACAAAGA 1927
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   1861 TGTTGGACGTGGTCACGAGACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACAAAGA 1920

Qy   1928 ACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATG 1987
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 ACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATG 1980

Qy   1988 CTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGCCAAATGCAAGTC 2047
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 CTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGCCAAATGCAAGTC 2040

Qy   2048 TAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGG 2107
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 TAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGG 2100

Qy   2108 AAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATG 2167
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 AAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATG 2160

Qy   2168 AATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGA 2227
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 AATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGA 2220

Qy   2228 TAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGA 2287
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 TAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGA 2280

Qy   2288 TCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTCTCTTCCC 2347
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 TCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTCTCTTCCC 2340

Qy   2348 CACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACCAGAGAAG 2407
          ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   2341 CACCCTCTTATTTCCAGCAGACCCATATCCAACAAGACCCGGCACTGCCAACCAGAGAAG 2400

Qy   2408 GCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAAT 2467
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 GCAAAGAAGGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAAT 2460

Qy   2468 ATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGCAACTGCA 2527
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 ATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGCAACTGCA 2520

Qy   2528 GAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTCTCTGCGA 2587
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 GAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTCTCTGCGA 2580

Qy   2588 CCCTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTACCTACAATATG 2641
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 CCCTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTACCTACAATATG 2634
```

```
RESULT 11
SIVMM132ZO
LOCUS       SIVMM132ZO              2781 bp ss-DNA     linear   VRL 21-DEC-1993
DEFINITION  Simian immunodeficiency virus (Mm132zd) envelope glycoprotein and
            nef protein genes, 5'ends.
ACCESSION   L22812
VERSION     L22812.1  GI:431483
KEYWORDS    envelope glycoprotein; nef protein.
SOURCE      Simian immunodeficiency virus DNA.
  ORGANISM  Simian immunodeficiency virus
            Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
            lentivirus group.
REFERENCE   1  (bases 1 to 2781)
  AUTHORS   Kodama,T., Mori,K., Kawahara,T., Ringler,D.J. and Desrosiers,R.C.
  TITLE     Analysis of simian immunodeficiency virus sequence variation in
            tissues of rhesus macaques with simian AIDS
  JOURNAL   J. Virol. 67 (11), 6522-6534 (1993)
  MEDLINE   94016835
   PUBMED   8411355
FEATURES             Location/Qualifiers
     source          1. .2781
                     /organism="Simian immunodeficiency virus"
                     /proviral
                     /db_xref="taxon:11723"
     gene            1. .2453
                     /gene="tat"
     exon            1. .147
                     /gene="tat"
                     /number=1
     gene            79. .2610
                     /gene="rev"
     exon            79. .147
                     /gene="tat"
                     /number=1
     misc_signal     147
                     /gene="tat"
                     /note="splice donor site"
     gene            155. .2781
                     /gene="env"
     CDS             155. .>2781
                     /gene="env"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAA47618.1"
                     /db_xref="GI:431484"
                     /translation="MGCLGNQLLIAILLLSVYGIYCTLYVTVFYGVPAWRNATIPLFC
                     ATKNRDTWGTTQCLPDNGDYSEMALNVTESFDAWNNTVTEQAIEDVWQLFETSIKPCV
                     KLSPLCITMRCNKSETDRWGLTKSITTTASTTSTTASAKVDMVDETSSCIAQDNCAGL
                     EQEQMISCKFNMTGLERDKKKEYNETWYSADLVCEQGNNTGNESRCYMNHCNTSVIQE
                     SCDKHYWDAIRFRYCAPPGYALLRCNDTNYSGFMPKCSKVVVSSCTRMMETQASTWFG
                     FNGTRAENRTYIYWHGRDNRTIISLNKYYNLTMKCRRPGNKTVSPVTIMSGLVFHSQP
                     INDRPKQAWCRFGGKWKDAIKEVKQTIVKHPRYTGTNNTDKINLTAPRGGNPEVTFMW
                     TNCRGEFLYCKMNWFLNWVEDRNTANQKPKEQHKRNYVPCHIRQIINTWHKVGKNVYL
                     PPREGDLTCNSTVTSLIANIDWIDGNQTNITMSAEVAGLYRLELGDYKLVEITPIGLA
                     PTDVKRYTTGGTSRNKRGVFVLGFLGFLATAGSAMGAASLTLTAQSRTLLAGIVQQQQ
                     QLLDVVKRQQELLRLTVWGTKNLQTRVTAIEKYLKDQAQLNAWGCAFRQVCHTTVPWP
                     NASLTPKWNNETWQEWERKVDFLEENITALLEEAQIQQEKNMYELQKLNSWDVFGNWF
                     DLSSWIKYIQYGVYIVVGVILLRIVIYIVQMLAKLRQGYRPVFSSPPSYFQQTHIQQD
                     PALPTREGKERDGGEGGGNSSWPWQIEYIHFLIRQLIRLLTWLFSNCRTLLSRVYQIL
                     QPILQRLSATLQRIREVLRTELTYLQYGWSYFHEAVQAVWRSATETLAGAWGNLWETL
                     RRGGRGILAIPRRIRQGLEL"
     misc_signal     1731
                     /gene="tat"
                     /note="SU/TM cleavage site"
     exon            2354. .2610
```

```
                     /gene="rev"
                     /number=2
     exon            2354. .2453
                     /gene="tat"
                     /number=2
     misc_signal     2355
                     /gene="tat"
                     /note="splice acceptor"
     gene            2628. .2781
                     /gene="nef"
     CDS             2628. .2781
                     /partial
                     /gene="nef"
                     /codon_start=1
                     /product="nef protein"
                     /protein_id="AAA47619.1"
                     /db_xref="GI:555343"
                     /translation="MGGAISMRRSRPSGDLRQRLLRARGETYGRLLGEVEEGYSQSPG
                     GLDKGLS"
BASE COUNT      896 a    523 c    669 g    693 t
ORIGIN

  Query Match             98.8%;  Score 2609.2;  DB 14;  Length 2781;
  Best Local Similarity   99.5%;  Pred. No. 0;
  Matches 2617;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy      12 GCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTGTTTTCTTAAAA 71
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTGTTTTCTTAAAA 60

Qy      72 AAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCCGAAAAAGGCTA 131
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCCGAAAAAGGCTA 120

Qy     132 AGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGGAATCAGCTGCT 191
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 AGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGGAATCAGCTGCT 180

Qy     192 TATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATATGTCACAGTCTT 251
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATATGTCACAGTCTT 240

Qy     252 TTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCAACCAAGAATAG 311
           |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TTATGGTGTACCGGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCAACCAAGAATAG 300

Qy     312 GGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCAGAAGTGGCCCT 371
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||| |||||
Db     301 GGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCAGAAATGGCCCT 360

Qy     372 TAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAGGCAATAGAGGA 431
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAGGCAATAGAGGA 420

Qy     432 TGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCCCCATTATGCAT 491
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCCCCATTATGCAT 480

Qy     492 TACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAATCAATAACAAC 551
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 TACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAATCAATAACAAC 540

Qy     552 AACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTCAATGAGACTAG 611
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db     541 AACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTCGATGAGACTAG 600

Qy     612 TTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCAAATGATAAGCTGTAA 671
           |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db     601 TTCTTGTATAGCCCAGGATAATTGCGCAGGCTTGGAACAAGAGCAAATGATAAGCTGTAA 660

Qy     672 ATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAATGAAACTTGGTACTC 731
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     661 ATTCAACATGACAGGGTTAGAAAGAGACAAGAAAAAAGAGTACAATGAAACTTGGTACTC 720

Qy     732 TGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGATGTTACATGAA 791
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 TGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGATGTTACATGAA 780

Qy     792 CCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGGGATGCTATTAG 851
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 CCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGGGATGCTATTAG 840

Qy     852 ATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGACACAAATTATTC 911
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 ATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGACACAAATTATTC 900

Qy     912 AGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGGATGATGGAGAC 971
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 AGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGGATGATGGAGAC 960

Qy     972 ACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGAACTTATATTTA 1031
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 ACAGGCTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGAACTTATATTTA 1020

Qy    1032 CTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTATAATCTAACAAT 1091
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 CTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTATAATCTAACAAT 1080

Qy    1092 GAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATGTCTGGATTGGT 1151
           |||||||||||||||||||||||||||||| |||||||||||||||| |||||||||||
Db    1081 GAAATGTAGAAGACCAGGAAATAAGACAGTTTCACCAGTCACCATTATGTCCGGATTGGT 1140

Qy    1152 TTTCCACTCACAACCAATCAATGATAGGCCAAAGCAGGCATGGTGTTGGTTTGGAGGAAA 1211
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    1141 TTTCCACTCACAACCAATCAATGATAGGCCAAAGCAGGCATGGTGTAGGTTTGGAGGAAA 1200

Qy    1212 ATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCCAGGTATACTGG 1271
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 ATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCCAGGTATACTGG 1260

Qy    1272 AACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGAGGAGGAGATCCGGAAGTTAC 1331
           |||||||||||||||||||||||||||||||| ||||||| ||||||||||||
Db    1261 AACTAACAATACTGATAAAATCAATTTGACGGCTCCTAGAGGAGGAAATCCGGAAGTTAC 1320

Qy    1332 CTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATGAATTGGTTTCTAAA 1391
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 CTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATGAATTGGTTTCTAAA 1380

Qy    1392 TTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGGAACAGCATAAAAGGAATTA 1451
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 TTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGGAACAGCATAAAAGGAATTA 1440

Qy    1452 CGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAGGCAAAAATGTTTA 1511
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 CGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAGGCAAAAATGTTTA 1500

Qy    1512 TTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACCAGTCTCATAGCAAA 1571
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 TTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACCAGTCTCATAGCAAA 1560

Qy    1572 CATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTGCAGAGGTGGCAGAACT 1631
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    1561 CATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTGCAGAGGTGGCAGGACT 1620

Qy    1632 GTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTCCAATTGGCTTGGCCCC 1691
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 GTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTCCAATTGGCTTGGCCCC 1680

Qy    1692 CACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCTTTGT 1751
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1681 CACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCTTTGT 1740

Qy    1752 GCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGTTGAC 1811
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1741 GCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGTTGAC 1800
```

```
Qy   1812 GCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGCTGTT 1871
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 GCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGCTGTT 1860

Qy   1872 GGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACAAAGAACCT 1931
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 GGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACAAAGAACCT 1920

Qy   1932 CCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATGCTTG 1991
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 CCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATGCTTG 1980

Qy   1992 GGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGCCAAATGCAAGTCTAAC 2051
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 GGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGCCAAATGCAAGTCTAAC 2040

Qy   2052 ACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGGAAGA 2111
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 ACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGGAAGA 2100

Qy   2112 AAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATGAATT 2171
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 AAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATGAATT 2160

Qy   2172 ACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGATAAA 2231
          ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   2161 ACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTTCTTCTTGGATAAA 2220

Qy   2232 GTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGATCTA 2291
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 GTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGATCTA 2280

Qy   2292 TATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTCTCTTCCCCACC 2351
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 TATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTCTCTTCCCCACC 2340

Qy   2352 CTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACCAGAGAAGGCAA 2411
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 CTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACCAGAGAAGGCAA 2400

Qy   2412 AGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAATATAT 2471
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 AGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAATATAT 2460

Qy   2472 TCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGCAACTGCAGAAC 2531
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 TCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGCAACTGCAGAAC 2520

Qy   2532 CTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTCTCTGCGACCCT 2591
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 CTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTCTCTGCGACCCT 2580

Qy   2592 ACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTACCTACAATATG 2641
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 ACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTACCTACAATATG 2630
```

```
RESULT 12
AY033146
LOCUS       AY033146               10279 bp    DNA     linear   VRL 05-JUL-2001
DEFINITION  Simian immunodeficiency virus clone 17E-Fr, complete sequence.
ACCESSION   AY033146
VERSION     AY033146.1  GI:14600336
KEYWORDS    .
SOURCE      Simian immunodeficiency virus 17E-Fr.
  ORGANISM  Simian immunodeficiency virus 17E-Fr
            Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
            lentivirus group.
REFERENCE   1  (bases 1 to 10279)
  AUTHORS   Flaherty,M.T., Hauer,D.A., Mankowski,J.L., Zink,M.C. and
            Clements,J.E.
  TITLE     Molecular and biological characterization of a neurovirulent
            molecular clone of simian immunodeficiency virus
  JOURNAL   J. Virol. 71 (8), 5790-5798 (1997)
  MEDLINE   97366637
   PUBMED   9223467
REFERENCE   2  (bases 1 to 10279)
  AUTHORS   Flaherty,M.T., Hauer,D.A., Mankowski,J.L., Zink,M.C. and
            Clements,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-APR-2001) Comp. Med., Johns Hopkins U. Sch. of Med.,
            600 N. Wolfe St., Baltimore, MD 21287, USA
FEATURES             Location/Qualifiers
     source          1. .10279
                     /organism="Simian immunodeficiency virus 17E-Fr"
                     /proviral
                     /db_xref="taxon:160753"
                     /clone="17E-Fr"
                     /note="recombinant of SIV/17E-Br and SIVmac239; envelope,
                     nef, and 3' LTR from SIV/17E-Br swarm were cloned in place
                     of the same genes in an infectious molecular clone of
                     SIVmac239"
     LTR             1. .818
                     /note="5' long terminal repeat"
     misc_feature    1. .517
                     /note="U3 region"
     misc_feature    518. .694
                     /note="R region"
     misc_feature    695. .818
                     /note="U5 region"
     gene            1053. .2585
                     /gene="gag"
     CDS             1053. .2585
                     /gene="gag"
                     /codon_start=1
                     /product="gag polyprotein"
                     /protein_id="AAK64193.1"
                     /db_xref="GI:14600340"
                     /translation="MGVRNSVLSGKKADELEKIRLRPNGKKKYMLKHVVWAANELDRF
                     GLAESLLENKEGCQKILSVLAPLVPTGSENLKSLYNTVCVIWCIHAEEKVKHTEEAKQ
                     IVQRHLVVETGTTETMPKTSRPTAPSSGRGGNYPVQQIGGNYVHLPLSPRTLNAWVKL
                     IEEKKFGAEVVPGFQALSEGCTPYDINQMLNCVGDHQAAMQIIRDIINEEAADWDLQH
                     PQPAPQQGQLREPSGSDIAGTTSSVDEQIQWMYRQQNPIPVGNIYRRWIQLGLQKCVR
                     MYNPTNILDVKQGPKEPFQSYVDRFYKSLRAEQTDAAVKNWMTQTLLIQNANPDCKLV
                     LKGLGVNPTLEEMLTACQGVGGPGQKARLMAEALKEALAPVPIPFAAAQQRGPRKPIK
                     CWNCGKEGHSARQCRAPRRQGCWKCGKMDHVMAKCPDRQAGFLGLGPWGKKPRNFPMA
                     QVHQGLMPTAPPEDPAVDLLKNYMQLGKQQREKQRESREKPYKEVTEDLLHLNSLFGG
                     DQ"
     gene            <2228. .5410
                     /gene="pol"
     CDS             <2228. .5410
                     /gene="pol"
                     /codon_start=1
                     /product="pol polyprotein"
                     /protein_id="AAK64194.1"
                     /db_xref="GI:14600341"
                     /translation="VLELWERGTLCKAMQSPKKTGMLEMWKNGPCYGQMPRQTGGFFR
                     PWSMGKEAPQFPHGSSASGADANCSPRGPSCGSAKELHAVGQAAERKAERKQREALQG
                     GDRGFAAPQFSLWRRPVVTAHIEGQPVEVLLDTGADDSIVTGIELGPHYTPKIVGGIG
                     GFINTKEYKNVEIEVLGKRIKGTIMTGDTPINIFGRNLLTALGMSLNFPIAKVEPVKV
                     ALKPGKDGPKLKQWPLSKEKIVALREICEKMEKDGQLEEAPPTNPYNTPTFAIKKKDK
                     NKWRMLIDFRELNRVTQDFTEVQLGIPHPAGLAKRKRITVLDIGDAYFSIPLDEEFRQ
                     YTAFTLPSVNNAEPGKRYIYKVLPQGWKGSPAIFQYTMRHVLEPFRKANPDVTLVQYM
                     DDILIASDRTDLEHDRVVLQSKELLNSIGFSTPEEKFQKDPPFQWMGYELWPTKWKLQ
                     KIELPQRETWTVNDIQKLVGVLNWAAQIYPGIKTKHLCRLIRGKMTLTEEVQWTEMAE
                     AEYEENKIILSQEQEGCYYQEGKPLEATVIKSQDNQWSYKIHQEDKILKVGKFAKIKN
                     THTNGVRLLAHVIQKIGKEAIVIWGQVPKFHLPVEKDVWEQWWTDYWQVTWIPEWDFI
                     STPPLVRLVFNLVKDPIEGEETYYTDGSCNKQSKEGKAGYITDRGKDKVKVLEQTTNQ
                     QAELEAFLMALTDSGPKANIIVDSQYVMGIITGCPTESESRLVNQIIEEMIKKSEIYV
                     AWVPAHKGIGGNQEIDHLVSQGIRQVLFLEKIEPAQEEHDKYHSNVKELVFKFGLPRI
                     VARQIVDTCDKCHQKGEAIHGQANSDLGTWQMDCTHLEGKIIIVAVHVASGFIEAEVI
                     PQETGRQTALFLLKLAGRWPITHLHTDNGANFASQEVKMVAWWAGIEHTFGVPYNPQS
                     QGVVEAMNHHLKNQIDRIREQANSVETIVLMAVHCMNFKRRGGIGDMTPAERLINMIT
                     TEQEIQFQQSKNSKFKNFRVYYREGRDQLWKGPGELLWKGEGAVILKVGTDIKVVPRR
                     KAKIIKDYGGGKEVDSSSHMEDTGEAREVA"
     misc_feature    5084. .5099
```

```
                     /gene="pol"
                     /note="polypurine tract"
     gene            5340. .5984
                     /gene="vif"
     CDS             5340. .5984
                     /gene="vif"
                     /codon_start=1
                     /product="vif protein"
                     /protein_id="AAK64195.1"
                     /db_xref="GI:14600342"
                     /translation="MEEEKRWIAVPTWRIPERLERWHSLIKYLKYKTKDLQKVCYVPH
                     FKVGWAWWTCSRVIFPLQEGSHLEVQGYWHLTPEKGWLSTYAVRITWYSKNFWTDVTP
                     NYADILLHSTYFPCFTAGEVRRAIRGEQLLSCCRFPRAHKYQVPSLQYLALKVVSDVR
                     SQGENPTWKQWRRDNRRGLRMAKQNSRGDKQRGGKPPTKGANFPGLAKVLGILA"
     gene            5812. .6150
                     /gene="vpx"
     CDS             5812. .6150
                     /gene="vpx"
                     /codon_start=1
                     /product="vpx protein"
                     /protein_id="AAK64196.1"
                     /db_xref="GI:14600343"
                     /translation="MSDPRERIPPGNSGEETIGEAFEWLNRTVEEINREAVNHLPREL
                     IFQVWQRSWEYWHDEQGMSPSYVKYRYLCLIQKALFMHCKKGCRCLGEGHGAGGWRPG
                     PPPPPPPGLA"
     gene            6151. .6456
                     /gene="vpr"
     CDS             6151. .6456
                     /gene="vpr"
                     /codon_start=1
                     /product="vpr protein"
                     /protein_id="AAK64197.1"
                     /db_xref="GI:14600344"
                     /translation="MEERPPENEGPQREPWDEWVVEVLEELKEEALKHFDPRLLTALG
                     NHIYNRHGDTLEGAGELIRILQRALFMHFRGGCIHSRIGQPGGGNPLSAIPPSRSML"
     gene            6302. .8902
                     /gene="tat"
     CDS             join(6302. .6597,8803. .8902)
                     /gene="tat"
                     /note="transactivator of viral transcription"
                     /codon_start=1
                     /product="tat protein"
                     /protein_id="AAK64190.1"
                     /db_xref="GI:14600337"
                     /translation="METPLREQENSLESSNERSSCISEADASTPESANLGEEILSQLY
                     RPLEACYNTCYCKKCCYHCQFCFLKKGLGICYEQSRKRRRTPKKAKANTSSASNNRPI
                     SNRTRHCQPEKAKKETVEKAVATAPGLDR"
     gene            6528. .9059
                     /gene="rev"
     CDS             join(6528. .6597,8803. .9059)
                     /gene="rev"
                     /codon_start=1
                     /product="rev protein"
                     /protein_id="AAK64191.1"
                     /db_xref="GI:14600338"
                     /translation="MNNHEREEELRKRLRLIHLLHQTTDPYPTGPGTANQRRQRRRRW
                     RRRWQQLLALTDRIYSFPDPPTDTPLDLAIQQLQNLAIESIPDPPTNTPEALCDPTED
                     SRSPQD"
     gene            6604. .9243
                     /gene="env"
                     /note="contains premature stop codon"
     CDS             6604. .8895
                     /gene="env"
                     /codon_start=1
                     /product="truncated envelope glycoprotein"
                     /protein_id="AAK64192.1"
                     /db_xref="GI:14600339"
                     /translation="MGCLGNQLLIAILLLSVYGIYCTLYVTVFYGVPVWRNATIPLFC
                     ATKNRDTWGTTQCLPDNGDYSEMALNVTESFDAWNNTVTEQAIEDVWQLFETSIKPCV
                     KLSPLCITMRCNKSETDRWGLTKSITTTASTTSTTASARVDMVNETSSCIAQDNCAGL
                     EQEQMISCKFNMTGLNRDKKKEYNETWYSADLVCEQGNNTGNESRCYMNHCNTSVIKE
                     SCDKHYWDAIRFRYCAPPGYALLRCNDTNYSGFMPKCSKVVVSSCTRMMETQTSTWFG
                     FNGTRAENRTYIYWHGRDNRTIISLNKYYNLTIKCRRPGNKTVLPVTIMSGLVFHSQR
                     INDRPRQAWCWFGGKWKDAIKEVKQTIVKHPRYTGTNNTDKINLTAPRGGDPEVTFMW
                     TNCRGEFLYCKMNWFLNWVEDRNTANQKPKEQHKRNYVPCHIRQIINTWHKVGKNVYL
                     PPREGDLTCNSTVTSLIANIDWIDGNQTNITMSAEVAELYRLELGDYKLVEITPIGLA
                     PTDVKRYTTGGTSRNKRGVFVLGFLGFLATAGSAMGAASLTLTAQSRTLLAGIVQQQQ
                     QLLDVVKRQQELLRLTVWGTKNLQTRVTAIEKYLKDQAQLNTWGCAFRQVCHTTVPWP
                     NASLIPKWNNETWQEWERKVDFLEENITALLEEAQIQQEKNMYELQKLNSWNVFGNWF
                     DLASWIKYIQYGVYIVVGVILLRIVIYIVQVLAKLRQGYRPVFSSPPSYFQQTHIQQD
                     PALPTREGKEGDGGEGGGNSSWP"
     gene            9077. .9868
                     /gene="nef"
     CDS             9077. .9868
                     /gene="nef"
                     /codon_start=1
                     /product="nef protein"
                     /protein_id="AAK64198.1"
                     /db_xref="GI:14600345"
                     /translation="MGGAISMRRSRSSGDLRQRLLRARGETYGRLLGEVEDGYSQSPG
                     GLDKGLSSLSCEGQKYNQGQYMNTPWRNPAKEREKLAYRKQNMDDIDEQDDDLVGVSV
                     RPKVPLRTMSYKLAIDISHFIKEKGGLEGIYYSARRHRILDIYLEKEKGIIPDWQDYT
                     SGPGIRFPKTFGWLWKLVPVNVSDEAQEDEEHYLMHPAQTSQWDDPWGEVLAWKFDPT
                     LAYTYEAYVRYPEEFGSKSGLSEEEVRRRLTARGLLNMADKKETR"
     misc_feature    9445. .9460
                     /gene="nef"
                     /note="polypurine tract"
     LTR             9462. .10279
                     /note="3' long terminal repeat"
     misc_feature    9462. .9978
                     /note="U3 region"
     misc_feature    9979. .10155
                     /note="R region"
     polyA_signal    10132. .10137
     misc_feature    10156. .10279
                     /note="U5 region"
BASE COUNT     3470 a   1932 c   2566 g   2311 t
ORIGIN

  Query Match             98.8%;  Score 2609;  DB 14;  Length 10279;
  Best Local Similarity   99.2%;  Pred. No. 0;
  Matches 2621;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy       1 TCTAGAAGCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6439 TCTAGAAGCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTG 6498

Qy      61 TTTTCTTAAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCC 120
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    6499 TTTTCTTAAAAAAGGCTTGGGGATATGTTATGAACAATCACGAAAGAGAAGAAGAACTCC 6558

Qy     121 GAAAAAGGCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6559 GAAAAAGGCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGG 6618

Qy     181 AATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6619 AATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATAT 6678

Qy     241 GTCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCA 300
          |||||||||||||||||||||| |||||||||||||||||||||| ||||||||
Db    6679 GTCACAGTCTTTTATGGTGTACCAGTTTGGAGGAATGCGACAATTCCCCTTTTTTGTGCA 6738

Qy     301 ACCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6739 ACCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCA 6798

Qy     361 GAAGTGGCCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAG 420
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6799 GAAATGGCCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAG 6858

Qy     421 GCAATAGAGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6859 GCAATAGAGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCC 6918

Qy     481 CCATTATGCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAA 540
```

```
Db   6919 CCATTATGCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAA 6978

Qy    541 TCAATAACAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTC 600

Db   6979 TCAATAACAACAACAGCATCAACAACATCAACGACAGCATCAGCAAGAGTAGACATGGTC 7038

Qy    601 AATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCAAATG 660

Db   7039 AATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCGCAGGCTTGGAACAAGAGCAAATG 7098

Qy    661 ATAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAATGAA 720

Db   7099 ATAAGCTGTAAATTCAACATGACAGGGTTAAATAGAGACAAGAAAAAAGAGTACAATGAA 7158

Qy    721 ACTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGA 780

Db   7159 ACCTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGA 7218

Qy    781 TGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGG 840

Db   7219 TGTTACATGAACCACTGTAACACTTCTGTTATCAAAGAGTCTTGTGACAAACATTATTGG 7278

Qy    841 GATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGAC 900

Db   7279 GATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGAC 7338

Qy    901 ACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGG 960

Db   7339 ACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGG 7398

Qy    961 ATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGA 1020

Db   7399 ATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGA 7458

Qy   1021 ACTTATATTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTAT 1080

Db   7459 ACTTATATTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTAT 7518

Qy   1081 AATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATG 1140

Db   7519 AATCTAACAATAAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATG 7578

Qy   1141 TCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGGCCAAAGCAGGCATGGTGTTGG 1200

Db   7579 TCTGGATTGGTTTTCCACTCACAACGAATCAATGATAGGCCAAGGCAGGCATGGTGTTGG 7638

Qy   1201 TTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCC 1260

Db   7639 TTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCC 7698

Qy   1261 AGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGAGGAGGAGAT 1320

Db   7699 AGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTAGAGGAGGAGAT 7758

Qy   1321 CCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATGAAT 1380

Db   7759 CCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATGAAT 7818

Qy   1381 TGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGGAACAGCAT 1440

Db   7819 TGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGGAACAGCAT 7878

Qy   1441 AAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAGGC 1500

Db   7879 AAAAGGAATTACGTGCCATGTCATATCAGACAAATAATCAACACTTGGCATAAAGTAGGC 7938

Qy   1501 AAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACCAGT 1560

Db   7939 AAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACCAGT 7998

Qy   1561 CTCATAGCAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTGCAGAG 1620

Db   7999 CTCATAGCAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTGCAGAG 8058

Qy   1621 GTGGCAGAACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTCCAATT 1680

Db   8059 GTGGCAGAACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTCCAATT 8118

Qy   1681 GGCTTGGCCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAAGA 1740

Db   8119 GGCTTGGCCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAAGA 8178

Qy   1741 GGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCG 1800

Db   8179 GGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCG 8238

Qy   1801 GCGTCGTTGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGCAACAG 1860

Db   8239 GCGTCGTTGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGCAACAG 8298

Qy   1861 CAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCTGGGGA 1920

Db   8299 CAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCTGGGGA 8358

Qy   1921 ACAAAGAACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAG 1980

Db   8359 ACAAAGAACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAG 8418

Qy   1981 CTGAATGCTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGCCAAAT 2040

Db   8419 CTGAATACTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGCCAAAT 8478

Qy   2041 GCAAGTCTAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGAC 2100

Db   8479 GCAAGTCTAATACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGAC 8538

Qy   2101 TTCTTGGAAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAAC 2160

Db   8539 TTCTTGGAAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAAC 8598

Qy   2161 ATGTATGAATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCT 2220

Db   8599 ATGTATGAATTACAAAAGTTGAATAGCTGGAATGTGTTTGGCAATTGGTTTGACCTTGCT 8658

Qy   2221 TCTTGGATAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGA 2280

Db   8659 TCTTGGATAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGA 8718

Qy   2281 ATAGTGATCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTC 2340

Db   8719 ATAGTGATCTATATAGTACAAGTGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTC 8778

Qy   2341 TCTTCCCCACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACC 2400

Db   8779 TCTTCCCCACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACC 8838

Qy   2401 AGAGAAGGCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAG 2460

Db   8839 AGAGAAGGCAAAGAAGGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGACAG 8898

Qy   2461 ATAGAATATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGC 2520

Db   8899 ATAGAATATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGC 8958

Qy   2521 AACTGCAGAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTC 2580

Db   8959 AACTGCAGAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTC 9018

Qy   2581 TCTGCGACCCTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTACCTACAATAT 2640

Db   9019 TCTGCGACCCTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTACCTACAATAT 9078

Qy   2641 G 2641

Db   9079 G 9079
```

```
RESULT 13
SIVMM316ZR
LOCUS       SIVMM316ZR              2799 bp ss-DNA     linear   VRL 21-DEC-1993
DEFINITION  Simian immunodeficiency virus (Mm316zf) envelope glycoprotein and
            nef protein genes, complete cds and 5'end.
ACCESSION   L22815
VERSION     L22815.1  GI:431576
KEYWORDS    envelope glycoprotein; nef protein.
SOURCE      Simian immunodeficiency virus DNA.
  ORGANISM  Simian immunodeficiency virus
            Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
            lentivirus group.
REFERENCE   1  (bases 1 to 2799)
  AUTHORS   Kodama,T., Mori,K., Kawahara,T., Ringler,D.J. and Desrosiers,R.C.
  TITLE     Analysis of simian immunodeficiency virus sequence variation in
            tissues of rhesus macaques with simian AIDS
  JOURNAL   J. Virol. 67 (11), 6522-6534 (1993)
  MEDLINE   94016835
   PUBMED   8411355
FEATURES             Location/Qualifiers
     source          1. .2799
                     /organism="Simian immunodeficiency virus"
                     /proviral
                     /db_xref="taxon:11723"
     gene            1. .2457
                     /gene="tat"
     exon            1. .150
                     /gene="tat"
                     /number=1
     gene            83. .2614
                     /gene="rev"
     exon            83. .150
                     /gene="tat"
                     /number=1
     misc_signal     150
                     /gene="tat"
                     /note="splice donor site"
     gene            159. .2798
                     /gene="env"
     CDS             159. .2798
                     /gene="env"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAA47694.1"
                     /db_xref="GI:431577"
                     /translation="MGCLGNQLLIAILLLSVYGIYCTLYVTVFYGVPVWRNATIPLFC
                     ATKNRDTWGTTQCLPDNGDYSEMALNVTESFDAWNNTVTEQAIEDVWQLFETSIKPCV
                     KLSPLCITMRCNKSETDKWGLTKSITTTASTTSTTASAKVDMVNETSSCIAQDNCAGL
                     EQEQMISCKFNMTGLKRNKKKEYNETWYSADLVCEQGNNTGNESRCYMNHCNTSVIQE
                     SCDKHYWDAIRFRYCAPPGYALLRCNDTNYSGFMPKCSKVVVSSCTRMMETQTSTWFG
                     FNGTRAENRTYIYWHGRDNRTIISLNKYYNLTMKCRRPGNKTVLPVTIMSGLVFHSQL
                     INDRPKQAWCWFGGKWKDAIKEVKQTIVKHPRYTGTNNTDKINLTAPRGGNPEATFMW
                     TNCRGEFLYCKMNWFLNWVEDRNTANQKSKEQHKRNYVPCHIRQIINTWHKVGKNVYL
                     PPREGDLTCNSTVTSLIANIDWIDGNQTNITMSAEVAELYRLELGDYKLVEITPIGLA
                     PTDVKRYTTGGTSRNKRGVFVLGFLGFLATAGSAMGAASLTLTAQSRTLLAGIVQQQQ
                     QLLDVVKRQQELLRLTVWGTKNLQTRVTAIERYLKDQAQLNAWGCAFRQVCHTTVPWP
                     NASLTPKWNNETWQEWERKVDFLEENITALLEEAQIQQEKNMYELQKLNSWDVFGNWF
                     DLASWIKYIQYGVYIVVGVILLRIVICIVQMLAKLRQGYRPVFSSPPSYFQQTHIQQD
                     PALPTREGKEGDGGEGGGNSSWPWQIKYIHFLIRQLIRLLTWLFSNCRTLLSRVYQIL
                     QPILQRLSATLQRIREVLRTELTYLQYGWSYFHEAVQAVWRSATETLAGAWGDLWETL
                     RRGGRWILAIPRRIRQGLELTLL"
     misc_signal     1734
                     /gene="tat"
                     /note="SU/TM cleavage site"
     exon            2358. .2614
                     /gene="rev"
                     /number=2
     exon            2358. .2457
                     /gene="tat"
                     /number=2
     misc_signal     2358
                     /gene="tat"
                     /note="splice acceptor"
     gene            2632. .2799
                     /gene="nef"
     CDS             2632. .2799
                     /partial
                     /gene="nef"
                     /codon_start=1
                     /product="nef protein"
                     /protein_id="AAA47695.1"
                     /db_xref="GI:431578"
                     /translation="MGGAISMRRSKPSGDLRQRLLRARGETYGRLLGEVEDGYSQSPG
                     GLDKGLSSLSCE"
BASE COUNT      905 a    524 c    666 g    704 t
ORIGIN
```

Query Match 98.8%; Score 2608.4; DB 14; Length 2799;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 2618; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

```
Qy       8 GCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTGTTTTCTT 67
Db       1 GCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTGTTTTCTT 60

Qy      68 AAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCCGAAAAAG 127
Db      61 AAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCCGAAAAAG 120

Qy     128 GCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGGAATCAGC 187
Db     121 GCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGGAATCAGC 180

Qy     188 TGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATATGTCACAG 247
Db     181 TGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATATGTCACAG 240

Qy     248 TCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCAACCAAGA 307
Db     241 TCTTTTATGGTGTACCAGTTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCAACCAAGA 300

Qy     308 ATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCAGAAGTGG 367
Db     301 ATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCAGAAATGG 360

Qy     368 CCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAGGCAATAG 427
Db     361 CCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAGGCAATAG 420

Qy     428 AGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCCCCATTAT 487
Db     421 AGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCCCCATTAT 480

Qy     488 GCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAATCAATAA 547
Db     481 GCATTACTATGAGATGCAATAAAAGTGAGACAGATAAATGGGGATTGACAAAATCAATAA 540

Qy     548 CAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTCAATGAGA 607
Db     541 CAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTCAATGAGA 600

Qy     608 CTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCAAATGATAAGCT 667
Db     601 CTAGTTCTTGTATAGCCCAGGATAATTGCGCAGGCTTGGAACAAGAGCAAATGATAAGCT 660

Qy     668 GTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAATGAAACTTGGT 727
Db     661 GTAAATTCAACATGACAGGGTTAAAAAGAAACAAGAAAAAAGAGTACAATGAAACTTGGT 720

Qy     728 ACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGATGTTACA 787
Db     721 ACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGATGTTACA 780

Qy     788 TGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGGGATGCTA 847
```

```
Db    781 TGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGGGATGCTA 840

Qy    848 TTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGACACAAATT 907
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 TTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGACACAAATT 900

Qy    908 ATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGGATGATGG 967
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 ATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGGATGATGG 960

Qy    968 AGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGAACTTATA 1027
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 AGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGAACTTATA 1020

Qy   1028 TTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTATAATCTAA 1087
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 TTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTATAATCTAA 1080

Qy   1088 CAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATGTCTGGAT 1147
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATGTCTGGAT 1140

Qy   1148 TGGTTTTCCACTCACAACCAATCAATGATAGGCCAAAGCAGGCATGGTGTTGGTTTGGAG 1207
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   1141 TGGTTTTCCACTCACAACTAATCAATGATAGGCCAAAGCAGGCATGGTGTTGGTTTGGAG 1200

Qy   1208 GAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCCAGGTATA 1267
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 GAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCCAGGTATA 1260

Qy   1268 CTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGAGGAGGAGATCCGGAAG 1327
          |||||||||||||||||||||||||||||||||| |||||||| |||||||||||||||
Db   1261 CTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTAGAGGAGGAAATCCGGAAG 1320

Qy   1328 TTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATGAATTGGTTTC 1387
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 CTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATGAATTGGTTTC 1380

Qy   1388 TAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGGAACAGCATAAAAGGA 1447
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   1381 TAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGTCAAAGGAACAGCATAAAAGGA 1440

Qy   1448 ATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAGGCAAAAATG 1507
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 ATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAGGCAAAAATG 1500

Qy   1508 TTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACCAGTCTCATAG 1567
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 TTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACCAGTCTCATAG 1560

Qy   1568 CAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTGCAGAGGTGGCAG 1627
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 CAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTGCAGAGGTGGCAG 1620

Qy   1628 AACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTCCAATTGGCTTGG 1687
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 AACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTCCAATTGGCTTGG 1680

Qy   1688 CCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCT 1747
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 CCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCT 1740

Qy   1748 TTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGT 1807
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 TTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGT 1800

Qy   1808 TGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGC 1867
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 TGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGC 1860

Qy   1868 TGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACAAAGA 1927
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1861 TGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACAAAGA 1920

Qy   1928 ACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATG 1987
          |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   1921 ACCTCCAGACTAGGGTCACTGCCATCGAGAGGTACTTAAAGGACCAGGCGCAGCTGAATG 1980

Qy   1988 CTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGCCAAATGCAAGTC 2047
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 CTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGCCAAATGCAAGTC 2040

Qy   2048 TAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGG 2107
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 TAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGG 2100

Qy   2108 AAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATG 2167
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 AAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATG 2160

Qy   2168 AATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGA 2227
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 AATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGA 2220

Qy   2228 TAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGA 2287
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 TAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGA 2280

Qy   2288 TCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTCTCTTCCC 2347
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 TCTGTATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTCTCTTCCC 2340

Qy   2348 CACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACCAGAGAAG 2407
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 CACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACCAGAGAAG 2400

Qy   2408 GCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAAT 2467
          |||||||| |||||||||||||||||||||||||||||||||||||||||| |||||||
Db   2401 GCAAAGAAGGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAAAAT 2460

Qy   2468 ATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGCAACTGCA 2527
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 ATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGCAACTGCA 2520

Qy   2528 GAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTCTCTGCGA 2587
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 GAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTCTCTGCGA 2580

Qy   2588 CCCTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTACCTACAATATG 2641
          ||||||| |||||||||||||||||||||||||||||| ||||||||||||||||
Db   2581 CCCTACAAAGGATTCGAGAAGTCCTCAGGACTGAACTAACCTACCTACAATATG 2634
```

```
RESULT 14
SIVMM132ZL
LOCUS       SIVMM132ZL              2781 bp ss-DNA     linear   VRL 21-DEC-1993
DEFINITION  Simian immunodeficiency virus (Mm132a) envelope glycoprotein and
            nef protein genes, 5'ends.
ACCESSION   L22809
VERSION     L22809.1  GI:431475
KEYWORDS    envelope glycoprotein; nef protein.
SOURCE      Simian immunodeficiency virus DNA.
  ORGANISM  Simian immunodeficiency virus
            Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
            lentivirus group.
REFERENCE   1  (bases 1 to 2781)
  AUTHORS   Kodama,T., Mori,K., Kawahara,T., Ringler,D.J. and Desrosiers,R.C.
  TITLE     Analysis of simian immunodeficiency virus sequence variation in
            tissues of rhesus macaques with simian AIDS
  JOURNAL   J. Virol. 67 (11), 6522-6534 (1993)
  MEDLINE   94016835
   PUBMED   8411355
FEATURES             Location/Qualifiers
     source          1. .2781
                     /organism="Simian immunodeficiency virus"
```

```
                     /proviral
                     /db_xref="taxon:11723"
     gene            1. .2453
                     /gene="tat"
     exon            1. .147
                     /gene="tat"
                     /number=1
     gene            79. .2610
                     /gene="rev"
     exon            79. .147
                     /gene="tat"
                     /number=1
     misc_signal     147
                     /gene="tat"
                     /note="splice donor site"
     CDS             155. .>2781
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAA47614.1"
                     /db_xref="GI:431476"
                     /translation="MGCLGNQLLIAILLLSVYGIYCTLYVTVFYGVPAWRNATIPLFC
                     AAKNRDTWGTTQCLPDNGDYSEMALNVTESFDAWNNTVTEQAIEDVWQLFETSIKPCV
                     KLSPLCITMRCNKSETDRWGLTKSITTTASTTSTTASAKVDMVDETSSCIAQDNCAGL
                     EQEQMISCKFNMTGLERDKKKEYNETWYSADLVCEQGNNTGNESRCYMNHCNTSVIQE
                     SCDKHYWDAIRFRYCAPPGYALLRCNDTNYSGFMPKCSKVVVSSCTRMMETQTSTWFG
                     FNGTRAENRTYIYWHGRDNRTIISLNKYYNLTMKCRRPGNKTVSPVTIMSGLVFHSQP
                     INDRPKQAWCRFGGKWKDAIKEVKQTIVKHPRYTGTNNTDKINLTAPRGGNPEVTFMW
                     TNCRGEFLYCKMNWFLNWVEDRNTANQKPKEQHKRNYVPCHIRQIINTWHKVGKNVYL
                     PPREGDLTCNSTVTSLIANIDWIDGNQTNITMSAEVAELYRLELGDYKLVEITPIGLA
                     PTDVKRYTTGGTSRNKRGVFVLGFLGFLATAGSAMGAASLTLTAQSRTLLAGIVQQQQ
                     QLLDVVKRQQELLRLTVWGTKNLQTRVTAIEKYLKDQAQLNAWGCAFRQVCHTTVPWP
                     NASLTPKWNNETWQEWERKVDFLEENITALLEEAQIQQEKNMYELQKLNSWDVFGNWF
                     DLASWIKYIQYGVYIVVGVILLRIVIYIVQMLAKLRQGYRPVFSSPPSYFQQTHIQQD
                     PALPTREGKEGDGGEGGGNSSWPWQIEYIHFLIRQLIRLLTWLFSNCRTLLSRVYQIL
                     QPILQRLSTTLQRIREVLRTELTYLQYGWSYFHEAVQAVWRSATETLAGAWGVLWETL
                     RRGGRWILAIPRRIRQGLEL"
     misc_signal     1731
                     /gene="tat"
                     /note="SU/TM cleavage site"
     exon            2354. .2610
                     /gene="rev"
                     /number=2
     exon            2354. .2453
                     /gene="tat"
                     /number=2
     misc_signal     2355
                     /gene="tat"
                     /note="splice acceptor"
     CDS             2628. .>2781
                     /codon_start=1
                     /product="nef protein"
                     /protein_id="AAA47615.1"
                     /db_xref="GI:555340" .
                     /translation="MGGAISMRRSRPSGDLRQRLLRARGESYGRLLGEVEDGYSQSPG
                     GLDKGLS"
BASE COUNT      895 a    523 c    669 g    694 t
ORIGIN

  Query Match             98.7%;  Score 2606;  DB 14;  Length 2781;
  Best Local Similarity   99.4%;  Pred. No. 0;
  Matches 2615;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy      12 GCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTGTTTTCTTAAAA 71
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTGTTTTCTTAAAA 60

Qy      72 AAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCCGAAAAAGGCTA 131
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCCGAAAAAGGCTA 120

Qy     132 AGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGGAATCAGCTGCT 191
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGGAATCAGCTGCT 180

Qy     192 TATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATATGTCACAGTCTT 251
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATATGTCACAGTCTT 240

Qy     252 TTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCAACCAAGAATAG 311
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCAGCCAAGAATAG 300

Qy     312 GGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCAGAAGTGGCCCT 371
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCAGAAATGGCCCT 360

Qy     372 TAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAGGCAATAGAGGA 431
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAGGCAATAGAGGA 420

Qy     432 TGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCCCCATTATGCAT 491
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCCCCATTATGCAT 480

Qy     492 TACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAATCAATAACAAC 551
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 TACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAATCAATAACAAC 540

Qy     552 AACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTCAATGAGACTAG 611
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 AACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTCGATGAGACTAG 600

Qy     612 TTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCAAATGATAAGCTGTAA 671
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 TTCTTGTATAGCCCAGGATAATTGCGCAGGCTTGGAACAAGAGCAAATGATAAGCTGTAA 660

Qy     672 ATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAATGAAACTTGGTACTC 731
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 ATTCAACATGACAGGGTTAGAAAGAGACAAGAAAAAAGAGTACAATGAAACTTGGTACTC 720

Qy     732 TGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGATGTTACATGAA 791
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 TGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGATGTTACATGAA 780

Qy     792 CCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGGGATGCTATTAG 851
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 CCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGGGATGCTATTAG 840

Qy     852 ATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGACACAAATTATTC 911
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 ATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGACACAAATTATTC 900

Qy     912 AGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGGATGATGGAGAC 971
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 AGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGGATGATGGAGAC 960

Qy     972 ACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGAACTTATATTTA 1031
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 ACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGAACTTATATTTA 1020

Qy    1032 CTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTATAATCTAACAAT 1091
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 CTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTATAATCTAACAAT 1080

Qy    1092 GAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATGTCTGGATTGGT 1151
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 GAAATGTAGAAGACCAGGAAATAAGACAGTTTCACCAGTCACCATTATGTCTGGATTGGT 1140

Qy    1152 TTTCCACTCACAACCAATCAATGATAGGCCAAAGCAGGCATGGTGTTGGTTTGGAGGAAA 1211
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 TTTCCACTCACAACCAATCAATGATAGGCCAAAGCAGGCATGGTGTAGGTTTGGAGGAAA 1200

Qy    1212 ATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCCAGGTATACTGG 1271
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 ATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCCAGGTATACTGG 1260
```

```
Qy   1272 AACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGAGGAGGAGATCCGGAAGTTAC 1331
          ||||||||||||||||||||||||||||||||||| |||||||| ||||||||||||
Db   1261 AACTAACAATACTGATAAAATCAATTTGACGGCTCCTAGAGGAGGAAATCCGGAAGTTAC 1320

Qy   1332 CTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATGAATTGGTTTCTAAA 1391
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 CTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATGAATTGGTTTCTAAA 1380

Qy   1392 TTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGGAACAGCATAAAAGGAATTA 1451
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 TTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGGAACAGCATAAAAGGAATTA 1440

Qy   1452 CGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAGGCAAAAATGTTTA 1511
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 CGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAGGCAAAAATGTTTA 1500

Qy   1512 TTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACCAGTCTCATAGCAAA 1571
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 TTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACCAGTCTCATAGCAAA 1560

Qy   1572 CATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTGCAGAGGTGGCAGAACT 1631
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 CATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTGCAGAGGTGGCAGAACT 1620

Qy   1632 GTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTCCAATTGGCTTGGCCCC 1691
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 GTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTCCAATTGGCTTGGCCCC 1680

Qy   1692 CACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCTTTGT 1751
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 CACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCTTTGT 1740

Qy   1752 GCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGTTGAC 1811
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 GCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGTTGAC 1800

Qy   1812 GCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGCTGTT 1871
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   1801 GCTGACCGCTCAGTCCCGAACCTTATTGGCTGGGATAGTGCAGCAACAGCAACAGCTGTT 1860

Qy   1872 GGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACAAAGAACCT 1931
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 GGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACAAAGAACCT 1920

Qy   1932 CCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATGCTTG 1991
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 CCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATGCTTG 1980

Qy   1992 GGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGCCAAATGCAAGTCTAAC 2051
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 GGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGCCAAATGCAAGTCTAAC 2040

Qy   2052 ACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGGAAGA 2111
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db   2041 ACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGGAGGA 2100

Qy   2112 AAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATGAATT 2171
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 AAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATGAATT 2160

Qy   2172 ACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGATAAA 2231
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 ACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGATAAA 2220

Qy   2232 GTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGATCTA 2291
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 GTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGATCTA 2280

Qy   2292 TATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTCTCTTCCCCACC 2351
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 TATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTCTCTTCCCCACC 2340

Qy   2352 CTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACCAGAGAAGGCAA 2411
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 CTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACCAGAGAAGGCAA 2400

Qy   2412 AGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAATATAT 2471
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 AGAAGGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAATATAT 2460

Qy   2472 TCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGCAACTGCAGAAC 2531
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 TCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGCAACTGCAGAAC 2520

Qy   2532 CTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTCTCTGCGACCCT 2591
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   2521 CTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTCTCTACGACCCT 2580

Qy   2592 ACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTACCTACAATATG 2641
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 ACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTACCTACAATATG 2630
```

RESULT 15
SIVMM316ZQ
LOCUS SIVMM316ZQ 2799 bp ss-DNA linear VRL 21-DEC-1993
DEFINITION Simian immunodeficiency virus (Mm316ze) envelope glycoprotein and nef protein genes, 5'ends.
ACCESSION L22814
VERSION L22814.1 GI:431574
KEYWORDS envelope glycoprotein; nef protein.
SOURCE Simian immunodeficiency virus DNA.
ORGANISM Simian immunodeficiency virus
Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate lentivirus group.
REFERENCE 1 (bases 1 to 2799)
AUTHORS Kodama,T., Mori,K., Kawahara,T., Ringler,D.J. and Desrosiers,R.C.
TITLE Analysis of simian immunodeficiency virus sequence variation in tissues of rhesus macaques with simian AIDS
JOURNAL J. Virol. 67 (11), 6522-6534 (1993)
MEDLINE 94016835
PUBMED 8411355

FEATURES	Location/Qualifiers
source	1. .2799
	/organism="Simian immunodeficiency virus"
	/proviral
	/db_xref="taxon:11723"
gene	1. .2457
	/gene="tat"
exon	1. .150
	/gene="tat"
	/number=1
gene	83. .2614
	/gene="rev"
exon	83. .150
	/gene="tat"
	/number=1
misc_signal	150
	/gene="tat"
	/note="splice donor site"
gene	159. .2798
	/gene="env"
CDS	159. .2798
	/gene="env"
	/note="internal stop codons at codon 2076 (TGA) and at codon 2703 (TGA)"
	/pseudo
	/codon_start=1
	/product="envelope glycoprotein"
misc_signal	1734
	/gene="tat"
	/note="SU/TM cleavage site"
exon	2358. .2614
	/gene="rev"
	/number=2
exon	2358. .2457

```
                     /gene="tat"
                     /number=2
     misc_signal     2358
                     /gene="tat"
                     /note="splice acceptor"
     gene            2632. .2799
                     /gene="nef"
     CDS             2632. .>2799
                     /gene="nef"
                     /codon_start=1
                     /product="nef protein"
                     /protein_id="AAA47693.1"
                     /db_xref="GI:431575"
                     /translation="MGGAISMRRSRPSGDLRQRLLRAREETYGRLLGEVEDGYSQSPG
                     GLDKGLSSLSCE"
BASE COUNT      901 a    526 c    669 g    703 t
ORIGIN

  Query Match             98.6%;  Score 2603.6;  DB 14;  Length 2799;
  Best Local Similarity   99.3%;  Pred. No. 0;
  Matches 2615;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy       8 GCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTGTTTTCTT 67
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTGTTTTCTT 60

Qy      68 AAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCCGAAAAAG 127
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCCGAAAAAG 120

Qy     128 GCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGGAATCAGC 187
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGGAATCAGC 180

Qy     188 TGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATATGTCACAG 247
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATATGTCACAG 240

Qy     248 TCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCAACCAAGA 307
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCAACCAAGA 300

Qy     308 ATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCAGAAGTGG 367
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db     301 ATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCAGAAATGG 360

Qy     368 CCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAGGCAATAG 427
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAGGCAATAG 420

Qy     428 AGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCCCCATTAT 487
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 AGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCCCCATTAT 480

Qy     488 GCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAATCAATAA 547
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 GCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAATCAATAA 540

Qy     548 CAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTCAATGAGA 607
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 CAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTCAATGAGA 600

Qy     608 CTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCAAATGATAAGCT 667
          |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db     601 CTAGTTCTTGTATAGCCCAGGATAATTGCGCAGGCTTGGAACAAGAGCAAATGATAAGCT 660

Qy     668 GTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAATGAAACTTGGT 727
          |||||||||||||||||||| ||||| ||||||||||||||||||||||||||||||||||
Db     661 GTAAATTCAACATGACAGGGTTAGAAAGAAACAAGAAAAAAGAGTACAATGAAACTTGGT 720

Qy     728 ACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGATGTTACA 787
          |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db     721 ACTCTGCAGATTTGGTATGTGGACAAGGGAATAACACTGGTAATGAAAGTAGATGTTACA 780

Qy     788 TGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGGGATGCTA 847
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 TGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGGGATGCTA 840

Qy     848 TTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGACACAAATT 907
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 TTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGACACAAATT 900

Qy     908 ATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGGATGATGG 967
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 ATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGGATGATGG 960

Qy     968 AGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGAACTTATA 1027
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 AGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGAACTTATA 1020

Qy    1028 TTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTATAATCTAA 1087
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    1021 TTTACTGGCATGGTAGGAATAATAGGACTATAATTAGTTTAAATAAGTATTATAATCTAA 1080

Qy    1088 CAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATGTCTGGAT 1147
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 CAATAAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATGTCTGGAT 1140

Qy    1148 TGGTTTTCCACTCACAACCAATCAATGATAGGCCAAAGCAGGCATGGTGTTGGTTTGGAG 1207
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 TGGTTTTCCACTCACAACCAATCAATGATAGGCCAAAGCAGGCATGGTGTTGGTTTGGAG 1200

Qy    1208 GAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCCAGGTATA 1267
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 GAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCCAGGTATA 1260

Qy    1268 CTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGAGGAGGAGATCCGGAAG 1327
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||| |||||||
Db    1261 CTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTAGAGGAGGAAATCCGGAAG 1320

Qy    1328 TTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATGAATTGGTTTC 1387
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 CTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATGAATTGGTTTC 1380

Qy    1388 TAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGGAACAGCATAAAAGGA 1447
          |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db    1381 TAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGTCAAAGGAACAGCATAAAAGGA 1440

Qy    1448 ATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAGGCAAAAATG 1507
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 ATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAGGCAAAAATG 1500

Qy    1508 TTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACCAGTCTCATAG 1567
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 TTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACCAGTCTCATAG 1560

Qy    1568 CAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTGCAGAGGTGGCAG 1627
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 CAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTGCAGAGGTGGCAG 1620

Qy    1628 AACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTCCAATTGGCTTGG 1687
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 AACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTCCAATTGGCTTGG 1680

Qy    1688 CCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCT 1747
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1681 CCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCT 1740

Qy    1748 TTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGT 1807
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1741 TTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGT 1800

Qy    1808 TGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGC 1867
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    1801 TGACGCTGACCGCTCAGTCCCGGACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGC 1860
```

```
Qy   1868 TGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACAAAGA 1927
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   1861 TGTTGGACGTGGTCACGAGACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACAAAGA 1920

Qy   1928 ACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATG 1987
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 ACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATG 1980

Qy   1988 CTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGCCAAATGCAAGTC 2047
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db   1981 CTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTATCATGGCCAAATGCAAGTC 2040

Qy   2048 TAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGG 2107
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   2041 TAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTAGGAGCGAAAGGTTGACTTCTTGG 2100

Qy   2108 AAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATG 2167
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 AGGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATG 2160

Qy   2168 AATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGA 2227
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   2161 AATTACAAAAGTTGAGTAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGA 2220

Qy   2228 TAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGA 2287
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db   2221 TAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTAA 2280

Qy   2288 TCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTCTCTTCCC 2347
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 TCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTCTCTTCCC 2340 -

Qy   2348 CACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACCAGAGAAG 2407
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 CACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACCAGAGAAG 2400

Qy   2408 GCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAAT 2467
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 GCAAAGAAGGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAAT 2460

Qy   2468 ATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGCAACTGCA 2527
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 ATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGCAACTGCA 2520

Qy   2528 GAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTCTCTGCGA 2587
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 GAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTCTCTGCGA 2580

Qy   2588 CCCTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTACCTACAATATG 2641
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 CCCTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTACCTACAATATG 2634
```

Search completed: November 16, 2002, 11:24:05
Job time : 4638.85 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 16, 2002, 06:33:28 ; Search time 363.611 Seconds
(without alignments)
16356.861 Million cell updates/sec

Title: M33262_COPY_6695_9335
Perfect score: 2641
Sequence: 1 TCTAGAAGCATGCTATAACA.....ACTGACCTACCTACAATATG 2641

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2641	100.0	10097	13	AAQ24802	SIVmac239 nef-dele
2	2641	100.0	10279	13	AAQ22487	SIVmac239 proviral
3	2641	100.0	10279	20	AAV81865	Nucleic acid seque
4	2538.6	96.1	10274	24	AAD25516	Simian-Human immun
5	2434.6	92.2	6978	24	AAD38657	CMVkan/R-R-SIVgp16
6	2434.6	92.2	7099	22	AAH22814	DNA sequence of a
7	2392.8	90.6	9601	9	AAN80860	Sequence corresp.
8	1745.4	66.1	2658	17	AAT09009	SIV subgenomic fra
9	1412.4	53.5	9643	9	AAN80859	Sequence of entire
10	1404.8	53.2	9672	13	AAQ20616	ROD HIV-2 isolate
11	1395.6	52.8	9360	11	AAQ02830	cDNA to HIV-2 RNA.
12	1377.2	52.1	9633	9	AAN80890	Sequence of cDNA c
13	1356.4	51.4	9473	10	AAN92768	HIV-2 variant HIV-
14	1356.4	51.4	9693	10	AAN92119	Sequence of clone
15	1345.8	51.0	9726	21	AAC81937	Packaging vector p
16	1345.6	51.0	9663	21	AAC81938	Packaging vector p
17	1344.8	50.9	10163	18	AAT61085	Full-length HIV-2K
18	1320.8	50.0	9630	21	AAZ89628	HIV-2 ROD DNA. Hu
19	1315	49.8	9822	10	AAN91774	Entire HIV-2/ST pr
20	1305.4	49.4	2673	9	AAN82199	ENVRN sequence fro
21	1303.8	49.4	2673	24	ABK52308	DNA encoding retro
22	1279.8	48.5	2673	24	ABK52305	cDNA encoding huma
23	1203.8	45.6	4204	21	AAC81925	Vector pCM-ENV(ROD
24	1181.4	44.7	4204	21	AAC81939	Envelope vector pC
25	949.2	35.9	1445	21	AAA29552	HIV codon altered
26	705.4	26.7	9465	21	AAA51008	Simian immunodefic
27	612	23.2	4414	10	AAN92771	HIV-2 variant HIV-
28	612	23.2	4414	16	AAQ92470	HIV-D205 partial s
29	601.2	22.8	1098	12	AAQ11965	Encodes 40kD "K3"
30	593.4	22.5	9170	22	AAF27365	SIV-derived retrov
31	590.8	22.4	9170	11	AAQ02829	DNA complementary
32	574	21.7	1445	21	AAA29551	HIV codon altered
33	515.4	19.5	792	21	AAC81942	HIV-2(ROD) RRE1 DN
34	514.4	19.5	1200	21	AAA29553	HIV codon altered
35	511.6	19.4	1440	12	AAQ11966	Encodes chimeric H
36	511.6	19.4	1797	12	AAQ11968	HIV-2/HIV-1/HTLV-I
37	510	19.3	906	12	AAQ11964	Encodes peptide fr
38	388.4	14.7	1109	17	AAT32686	HIV-2 transmembran
39	388.4	14.7	1119	17	AAT32684	HIV-2 transmembran
40	386.8	14.6	1119	11	AAQ04471	Synthetic gene enc
41	359.2	13.6	529	21	AAC81940	HIV-2(ROD) RRE2 DN
42	305.4	11.6	433	17	AAT28644	DNA fragment HE 20
43	286	10.8	5428	22	AAH77485	Mandrill immunodef
44	284.4	10.8	9641	22	AAH77486	Mandrill immunodef
45	275.2	10.4	9215	12	AAQ10203	Sequence of simian

ALIGNMENTS

RESULT 1
AAQ24802
ID AAQ24802 standard; DNA; 10097 BP.
XX
AC AAQ24802;
XX
DT 06-JUL-1992 (first entry)
XX
DE SIVmac239 nef-deletion.
XX
KW Macaque; monkey; polymerase chain reaction;
KW PCR; site-directed mutagenesis; retrovirus; null mutation; ss.
XX
OS Simian immunodeficiency virus.
XX
FH Key Location/Qualifiers
FT repeat_region 1..818
FT /*tag= a
FT /rpt_type= TERMINAL
FT /note= "i.e. 5' LTR"
FT repeat_unit 1..517
FT /*tag= b
FT /rpt_type= OTHER
FT /note= "U3"
FT repeat_unit 518..600
FT /*tag= c
FT /rpt_type= OTHER
FT /note= "R"
FT repeat_unit 601..818
FT /*tag= d
FT /rpt_type= OTHER
FT /note= "U5"

```
FT   primer_bind     822..849
FT                   /*tag= e
FT                   /standard_name= tRNA_PBS
FT   CDS             1053..2585
FT                   /*tag= f
FT                   /product= gag
FT   CDS             2228..5410
FT                   /*tag= g
FT                   /product= pol
FT   CDS             5340..5984
FT                   /*tag= h
FT                   /product= vif
FT   CDS             5812..6150
FT                   /*tag= i
FT                   /product= vpx
FT   CDS             6051..6456
FT                   /*tag= j
FT                   /product= vpr
FT   exon            6302..6597
FT                   /*tag= k
FT                   /product= tat
FT                   /note= "full-length product obtained by splicing"
FT   exon            6528..6597
FT                   /*tag= l
FT                   /product= rev
FT                   /note= "full-length product obtained by splicing"
FT   CDS             6604..9243
FT                   /*tag= m
FT                   /product= env
FT   exon            8803..8902
FT                   /*tag= n
FT                   /product= tat
FT                   /note= "see above"
FT   exon            8803..9059
FT                   /*tag= o
FT                   /product= rev
FT                   /note= "see above"
FT   CDS             9077..9686
FT                   /*tag= p
FT                   /product= nef
FT   repeat_region   9280..10097
FT                   /*tag= q
FT                   /rpt_type= TERMINAL
FT                   /note= "i.e. 3' LTR"
FT   misc_signal     415..424
FT                   /*tag= r
FT                   /standard_name= NF_Kappa_B
FT   GC_signal       429..438
FT                   /*tag= s
FT                   /standard_name= Sp1_binding_site
FT   GC_signal       440..449
FT                   /*tag= t
FT                   /standard_name= Sp1_binding_site
FT   GC_signal       451..460
FT                   /*tag= u
FT                   /standard_name= Sp1_binding_site
FT   GC_signal       462..471
FT                   /*tag= v
FT                   /standard_name= Sp1_binding_site
FT   TATA_signal     488..494
FT                   /*tag= w
FT   polyA_signal    9950..9955
FT                   /*tag= x
XX
PN   WO9200987-A.
XX
PD   23-JAN-1992.
XX
PF   10-JUL-1991;   91WO-US04884.
XX
PR   12-JUL-1990;   90US-0551945.
XX
PA   (HARD ) HARVARD COLLEGE.
XX
PI   Desrosiers RC;
XX
DR   WPI; 1992-056816/07.
DR   P-PSDB; AAR22365-R22371, AAR24126-7.
XX
PT   Primate lentivirus vaccine protecting against AIDS - and primate
PT   lentiviruses and their DNA clones contg. null mutations, useful for
PT   producing vaccine
XX
PS   Claim 1; Fig 1; 51pp; English.
XX
CC   Parental virus SIVmac239 was isolated from a macaque monkey (see
CC   AAQ22487). An SstI fragment of p239SpE3' contg. the C-terminus of
CC   the nef gene was cloned into M13 and subjected to site-directed
CC   mutagenesis. A 73mer primer was used which was complementary to
CC   bases 9215 through 9250 and 9433 through 9469. Since bases 9251
CC   through 9432 were not included in the primer, their complement was
CC   not included in the newly synthesised negative strand. Successful
CC   deletion was confirmed by DNA sequencing. The fragment contg. the
CC   deletion was cloned into p239SpE3' to yield p239S-E3'(nef-deletion)
CC   which was digested with SphI and ligated to SphI-cut p239SpSp5'.
CC   The ligation product was used to directly transfect cultured cells.
CC   See also AAQ21075-8.
XX
SQ   Sequence 10097 BP; 3387 A; 1911 C; 2527 G; 2272 T; 0 other;

  Query Match             100.0%;  Score 2641;  DB 13;  Length 10097;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2641;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TCTAGAAGCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6439 TCTAGAAGCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTG 6498

Qy      61 TTTTCTTAAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6499 TTTTCTTAAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCC 6558

Qy     121 GAAAAAGGCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6559 GAAAAAGGCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGG 6618

Qy     181 AATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6619 AATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATAT 6678

Qy     241 GTCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6679 GTCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCA 6738

Qy     301 ACCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6739 ACCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCA 6798

Qy     361 GAAGTGGCCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6799 GAAGTGGCCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAG 6858

Qy     421 GCAATAGAGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6859 GCAATAGAGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCC 6918

Qy     481 CCATTATGCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6919 CCATTATGCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAA 6978

Qy     541 TCAATAACAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6979 TCAATAACAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTC 7038

Qy     601 AATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCAAATG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   7039 AATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCAAATG 7098

Qy    661 ATAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAATGAA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7099 ATAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAATGAA 7158

Qy    721 ACTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7159 ACTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGA 7218

Qy    781 TGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7219 TGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGG 7278

Qy    841 GATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGAC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7279 GATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGAC 7338

Qy    901 ACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7339 ACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGG 7398

Qy    961 ATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGA 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7399 ATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGA 7458

Qy   1021 ACTTATATTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTAT 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7459 ACTTATATTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTAT 7518

Qy   1081 AATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATG 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7519 AATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATG 7578

Qy   1141 TCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGGCCAAAGCAGGCATGGTGTTGG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7579 TCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGGCCAAAGCAGGCATGGTGTTGG 7638

Qy   1201 TTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCC 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7639 TTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCC 7698

Qy   1261 AGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGAGGAGGAGAT 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7699 AGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGAGGAGGAGAT 7758

Qy   1321 CCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATGAAT 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7759 CCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATGAAT 7818

Qy   1381 TGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGGAACAGCAT 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7819 TGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGGAACAGCAT 7878

Qy   1441 AAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAGGC 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7879 AAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAGGC 7938

Qy   1501 AAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACCAGT 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7939 AAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACCAGT 7998

Qy   1561 CTCATAGCAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTGCAGAG 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7999 CTCATAGCAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTGCAGAG 8058

Qy   1621 GTGGCAGAACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTCCAATT 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8059 GTGGCAGAACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTCCAATT 8118

Qy   1681 GGCTTGGCCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAAGA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8119 GGCTTGGCCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAAGA 8178

Qy   1741 GGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCG 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8179 GGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCG 8238

Qy   1801 GCGTCGTTGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGCAACAG 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8239 GCGTCGTTGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGCAACAG 8298

Qy   1861 CAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCTGGGGA 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8299 CAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCTGGGGA 8358

Qy   1921 ACAAAGAACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAG 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8359 ACAAAGAACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAG 8418

Qy   1981 CTGAATGCTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGCCAAAT 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8419 CTGAATGCTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGCCAAAT 8478

Qy   2041 GCAAGTCTAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGAC 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8479 GCAAGTCTAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGAC 8538

Qy   2101 TTCTTGGAAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAAC 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8539 TTCTTGGAAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAAC 8598

Qy   2161 ATGTATGAATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCT 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8599 ATGTATGAATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCT 8658

Qy   2221 TCTTGGATAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGA 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8659 TCTTGGATAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGA 8718

Qy   2281 ATAGTGATCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTC 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8719 ATAGTGATCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTC 8778

Qy   2341 TCTTCCCCACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACC 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8779 TCTTCCCCACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACC 8838

Qy   2401 AGAGAAGGCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAG 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8839 AGAGAAGGCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAG 8898

Qy   2461 ATAGAATATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGC 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8899 ATAGAATATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGC 8958

Qy   2521 AACTGCAGAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTC 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8959 AACTGCAGAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTC 9018

Qy   2581 TCTGCGACCCTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTACCTACAATAT 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9019 TCTGCGACCCTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTACCTACAATAT 9078

Qy   2641 G 2641
          |
Db   9079 G 9079
```

RESULT 2
AAQ22487
ID AAQ22487 standard; DNA; 10279 BP.
XX
AC AAQ22487;
XX
DT 06-JUL-1992 (first entry)

```
XX
DE   SIVmac239 proviral genome.
XX
KW   Macaque; monkey; mac239; polymerase chain reaction;
KW   PCR; site-directed mutagenesis; retrovirus; ss.
XX
OS   Simian immunodeficiency virus.
XX
FH   Key             Location/Qualifiers
FT   repeat_region   1..818
FT                   /*tag= a
FT                   /rpt_type= TERMINAL
FT                   /note= "i.e. 5' LTR"
FT   repeat_unit     1..517
FT                   /*tag= b
FT                   /rpt_type= OTHER
FT                   /note= "U3"
FT   repeat_unit     518..600
FT                   /*tag= c
FT                   /rpt_type= OTHER
FT                   /note= "R"
FT   repeat_unit     601..818
FT                   /*tag= d
FT                   /rpt_type= OTHER
FT                   /note= "U5"
FT   primer_bind     822..849
FT                   /*tag= e
FT                   /standard_name= tRNA_PBS
FT   CDS             1053..2585
FT                   /*tag= f
FT                   /product= gag
FT   CDS             2228..5410
FT                   /*tag= g
FT                   /product= pol
FT   CDS             5340..5984
FT                   /*tag= h
FT                   /product= vif
FT   CDS             5812..6150
FT                   /*tag= i
FT                   /product= vpx
FT   CDS             6051..6456
FT                   /*tag= j
FT                   /product= vpr
FT   exon            6302..6597
FT                   /*tag= k
FT                   /product= tat
FT                   /note= "full-length product obtained by splicing"
FT   exon            6528..6597
FT                   /*tag= l
FT                   /product= rev
FT                   /note= "full-length product obtained by splicing"
FT   CDS             6604..9243
FT                   /*tag= m
FT                   /product= env
FT   exon            8803..8902
FT                   /*tag= n
FT                   /product= tat
FT                   /note= "see above"
FT   exon            8803..9059
FT                   /*tag= o
FT                   /product= rev
FT                   /note= "see above"
FT   CDS             9077..9868
FT                   /*tag= p
FT                   /product= nef
FT   repeat_region   9462..10279
FT                   /*tag= q
FT                   /rpt_type= TERMINAL
FT                   /note= "i.e. 3' LTR"
FT   misc_signal     415..424
FT                   /*tag= r
FT                   /standard_name= NF_Kappa_B
FT   GC_signal       429..438
FT                   /*tag= s
FT                   /standard_name= Sp1_binding_site
FT   GC_signal       440..449
FT                   /*tag= t
FT                   /standard_name= Sp1_binding_site
FT   GC_signal       451..460
FT                   /*tag= u
FT                   /standard_name= Sp1_binding_site
FT   GC_signal       462..471
FT                   /*tag= v
FT                   /standard_name= Sp1_binding_site
FT   TATA_signal     488..494
FT                   /*tag= w
FT   polyA_signal    10132..10137
FT                   /*tag= x
XX
PN   WO9200987-A.
XX
PD   23-JAN-1992.
XX
PF   10-JUL-1991;   91WO-US04884.
XX
PR   12-JUL-1990;   90US-0551945.
XX
PA   (HARD ) HARVARD COLLEGE.
XX
PI   Desrosiers RC;
XX
DR   WPI; 1992-056816/07.
DR   P-PSDB; AAR22365-R22372, AAR24126.
XX
PT   Primate lentivirus vaccine protecting against AIDS - and primate
PT   lentiviruses and their DNA clones contg. null mutations, useful for
PT   producing vaccine
XX
PS   Disclosure; Fig 1; 51pp; English.
XX
CC   Cell-free serum samples from a macaque monkey exhibiting symptoms
CC   characteristic of SIV infection were co-cultivated with HuT-78
CC   cells. Infectious SIVmac239 virus was identified in the cell
CC   supernatant. Total cell DNA was prepared from SIVmac239-infected
CC   cells and digested with EcoRI. An EMBL-4 library was constructed
CC   from 10-20kb EcoRI fragments (EcoRI is a non-cutter of SIVmac239).
CC   The library was screened with pK2 BamA as probe and a full-length
CC   molecular clone was isolated and sequenced. Then, EMBL-SIVmac239
CC   was digested with SphI and a 6706bp fragment, contg. the SphI site
CC   in the left flanking cellular DNA sequence to viral nucleotide no.
CC   6451, was inserted in vector pBS(+) to produce subclone p239SpSp5'.
CC   In a separate reaction, EMBL-SIVmac239 was digested with EcoRI and
CC   SphI. A 6361bp fragment from viral nucleotide 6452 to the EcoRI
CC   site in the right flanking cellular sequence, was inserted in
CC   pBS(-) to produce subclone p239SpE3'. These subclones were used to
CC   generate the full-length genomic sequence and to produce the
CC   preferred null-mutations of the invention. See Q24802 for
CC   nef-deletion mutant. See also Q21075-8.
XX
SQ   Sequence 10279 BP; 3465 A; 1936 C; 2569 G; 2309 T; 0 other;

  Query Match             100.0%;  Score 2641;  DB 13;  Length 10279;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2641;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TCTAGAAGCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6439 TCTAGAAGCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTG 6498

Qy      61 TTTTCTTAAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6499 TTTTCTTAAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCC 6558

Qy     121 GAAAAAGGCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6559 GAAAAAGGCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGG 6618
```

```
Qy    181 AATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6619 AATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATAT 6678

Qy    241 GTCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6679 GTCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCA 6738

Qy    301 ACCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6739 ACCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCA 6798

Qy    361 GAAGTGGCCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6799 GAAGTGGCCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAG 6858

Qy    421 GCAATAGAGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6859 GCAATAGAGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCC 6918

Qy    481 CCATTATGCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6919 CCATTATGCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAA 6978

Qy    541 TCAATAACAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6979 TCAATAACAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTC 7038

Qy    601 AATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCAAATG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7039 AATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCAAATG 7098

Qy    661 ATAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAATGAA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7099 ATAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAATGAA 7158

Qy    721 ACTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7159 ACTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGA 7218

Qy    781 TGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7219 TGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGG 7278

Qy    841 GATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGAC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7279 GATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGAC 7338

Qy    901 ACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7339 ACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGG 7398

Qy    961 ATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGA 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7399 ATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGA 7458

Qy   1021 ACTTATATTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTAT 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7459 ACTTATATTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTAT 7518

Qy   1081 AATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATG 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7519 AATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATG 7578

Qy   1141 TCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGGCCAAAGCAGGCATGGTGTTGG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7579 TCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGGCCAAAGCAGGCATGGTGTTGG 7638

Qy   1201 TTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCC 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7639 TTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCC 7698

Qy   1261 AGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGAGGAGGAGAT 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7699 AGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGAGGAGGAGAT 7758

Qy   1321 CCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATGAAT 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7759 CCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATGAAT 7818

Qy   1381 TGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGGAACAGCAT 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7819 TGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGGAACAGCAT 7878

Qy   1441 AAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAGGC 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7879 AAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAGGC 7938

Qy   1501 AAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACCAGT 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7939 AAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACCAGT 7998

Qy   1561 CTCATAGCAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTGCAGAG 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7999 CTCATAGCAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTGCAGAG 8058

Qy   1621 GTGGCAGAACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTCCAATT 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8059 GTGGCAGAACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTCCAATT 8118

Qy   1681 GGCTTGGCCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAAGA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8119 GGCTTGGCCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAAGA 8178

Qy   1741 GGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCG 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8179 GGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCG 8238

Qy   1801 GCGTCGTTGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGCAACAG 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8239 GCGTCGTTGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGCAACAG 8298

Qy   1861 CAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCTGGGGA 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8299 CAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCTGGGGA 8358

Qy   1921 ACAAAGAACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAG 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8359 ACAAAGAACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAG 8418

Qy   1981 CTGAATGCTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGCCAAAT 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8419 CTGAATGCTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGCCAAAT 8478

Qy   2041 GCAAGTCTAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGAC 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8479 GCAAGTCTAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGAC 8538

Qy   2101 TTCTTGGAAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAAC 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8539 TTCTTGGAAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAAC 8598

Qy   2161 ATGTATGAATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCT 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8599 ATGTATGAATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCT 8658

Qy   2221 TCTTGGATAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGA 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8659 TCTTGGATAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGA 8718

Qy   2281 ATAGTGATCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTC 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8719 ATAGTGATCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTC 8778

Qy   2341 TCTTCCCCACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACC 2400
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8779 TCTTCCCCACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACC 8838

Qy    2401 AGAGAAGGCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAG 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8839 AGAGAAGGCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAG 8898

Qy    2461 ATAGAATATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGC 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8899 ATAGAATATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGC 8958

Qy    2521 AACTGCAGAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTC 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8959 AACTGCAGAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTC 9018

Qy    2581 TCTGCGACCCTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTACCTACAATAT 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9019 TCTGCGACCCTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTACCTACAATAT 9078

Qy    2641 G 2641
          |
Db    9079 G 9079
```

RESULT 3
AAV81865
ID AAV81865 standard; DNA; 10279 BP.
XX
AC AAV81865;
XX
DT 01-JUN-1999 (first entry)
XX
DE Nucleic acid sequence of SIVmac239 genome.
XX
KW Antigenic composition; primate; lentivirus; nef gene; vaccine;
KW infection; AIDS; SIVmac239; ss.
XX
OS Simian immunodeficiency virus.
XX
PN US5851813-A.
XX
PD 22-DEC-1998.
XX
PF 27-JAN-1994; 94US-0188583.
XX
PR 27-JAN-1994; 94US-0188583.
PR 12-JUL-1990; 90US-0551945.
PR 09-JUL-1991; 91US-0727494.
XX
PA (HARD) HARVARD COLLEGE.
XX
PI Desrosiers RC;
XX
DR WPI; 1999-080408/07.
DR P-PSDB; AAW89313, AAW89314, AAW89315, AAW89316, AAW89317, AAW89318,
DR AAW89319, AAW89320, AAW89321.
XX
PT Lentivirus antigenic compositions - containing lentivirus with nef
PT gene deletion
XX
PS Disclosure; Fig 1A-V; 93pp; English.
XX
CC The invention relates to an antigenic composition comprising an isolated
CC primate lentivirus whose genome contains an engineered non-revertible
CC null mutation in the nef gene, or an infectious DNA clone in a carrier.
CC The antigenic composition is used in vaccines against infection by the
CC lentivirus, e.g. AIDS. The present sequence represents the nucleic
CC acid sequence of the SIVmac 239 genome.
XX
SQ Sequence 10279 BP; 3466 A; 1935 C; 2570 G; 2308 T; 0 other;

Query Match 100.0%; Score 2641; DB 20; Length 10279;
Best Local Similarity 100.0%; Pred. No. 0;

Matches 2641; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 TCTAGAAGCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6439 TCTAGAAGCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTG 6498

Qy      61 TTTTCTTAAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6499 TTTTCTTAAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCC 6558

Qy     121 GAAAAAGGCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6559 GAAAAAGGCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGG 6618

Qy     181 AATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6619 AATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATAT 6678

Qy     241 GTCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6679 GTCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCA 6738

Qy     301 ACCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6739 ACCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCA 6798

Qy     361 GAAGTGGCCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6799 GAAGTGGCCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAG 6858

Qy     421 GCAATAGAGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6859 GCAATAGAGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCC 6918

Qy     481 CCATTATGCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6919 CCATTATGCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAA 6978

Qy     541 TCAATAACAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6979 TCAATAACAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTC 7038

Qy     601 AATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCAAATG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7039 AATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCAAATG 7098

Qy     661 ATAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAATGAA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7099 ATAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAATGAA 7158

Qy     721 ACTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7159 ACTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGA 7218

Qy     781 TGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7219 TGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGG 7278

Qy     841 GATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGAC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7279 GATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGAC 7338

Qy     901 ACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7339 ACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGG 7398

Qy     961 ATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGA 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7399 ATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGA 7458

Qy    1021 ACTTATATTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTAT 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7459 ACTTATATTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTAT 7518
```

```
Qy   1081 AATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATG 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7519 AATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATG 7578

Qy   1141 TCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGGCCAAAGCAGGCATGGTGTTGG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7579 TCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGGCCAAAGCAGGCATGGTGTTGG 7638

Qy   1201 TTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCC 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7639 TTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCC 7698

Qy   1261 AGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGAGGAGGAGAT 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7699 AGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGAGGAGGAGAT 7758

Qy   1321 CCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATGAAT 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7759 CCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATGAAT 7818

Qy   1381 TGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGGAACAGCAT 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7819 TGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGGAACAGCAT 7878

Qy   1441 AAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAGGC 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7879 AAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAGGC 7938

Qy   1501 AAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACCAGT 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7939 AAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACCAGT 7998

Qy   1561 CTCATAGCAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTGCAGAG 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7999 CTCATAGCAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTGCAGAG 8058

Qy   1621 GTGGCAGAACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTCCAATT 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8059 GTGGCAGAACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTCCAATT 8118

Qy   1681 GGCTTGGCCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAAGA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8119 GGCTTGGCCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAAGA 8178

Qy   1741 GGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCG 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8179 GGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCG 8238

Qy   1801 GCGTCGTTGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGCAACAG 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8239 GCGTCGTTGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGCAACAG 8298

Qy   1861 CAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCTGGGGA 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8299 CAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCTGGGGA 8358

Qy   1921 ACAAAGAACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAG 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8359 ACAAAGAACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAG 8418

Qy   1981 CTGAATGCTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGCCAAAT 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8419 CTGAATGCTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGCCAAAT 8478

Qy   2041 GCAAGTCTAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGAC 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8479 GCAAGTCTAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGAC 8538

Qy   2101 TTCTTGGAAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAAC 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8539 TTCTTGGAAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAAC 8598
```

```
Qy   2161 ATGTATGAATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCT 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8599 ATGTATGAATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCT 8658

Qy   2221 TCTTGGATAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGA 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8659 TCTTGGATAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGA 8718

Qy   2281 ATAGTGATCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTC 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8719 ATAGTGATCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTC 8778

Qy   2341 TCTTCCCCACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACC 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8779 TCTTCCCCACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACC 8838

Qy   2401 AGAGAAGGCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAG 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8839 AGAGAAGGCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAG 8898

Qy   2461 ATAGAATATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGC 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8899 ATAGAATATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGC 8958

Qy   2521 AACTGCAGAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTC 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8959 AACTGCAGAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTC 9018

Qy   2581 TCTGCGACCCTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTACCTACAATAT 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9019 TCTGCGACCCTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTACCTACAATAT 9078

Qy   2641 G 2641
          |
Db   9079 G 9079
```

RESULT 4
AAD25516
ID AAD25516 standard; DNA; 10274 BP.
XX
AC AAD25516;
XX
DT 26-MAR-2002 (first entry)
XX
DE Simian-Human immunodeficiency virus (SHIV) clone 1A11 complete genome.
XX
KW HIV; human immunodeficiency virus; simian virus; cytostatic; cancer;
KW immunosuppressive; virucide; antibacterial; fungicide; protozoacide;
KW antirheumatic; antiinflammatory; antiarthritic; rheumatoid arthritis;
KW neuroprotective; multiple sclerosis; immune response; vasotropic;
KW vaccine; gene therapy; autoimmune disease; vasculitis; ds.
XX
OS Chimeric - Human immunodeficiency virus.
OS Chimeric - Simian virus.
XX
PN WO200176643-A1.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-US11372.
XX
PR 07-APR-2000; 2000US-195680P.
XX
PA (BAYU) BAYLOR COLLEGE MEDICINE.
XX
PI Orson FM, Kinsey BM, Bhogal BS;
XX
DR WPI; 2002-066308/09.
XX
PT Composition for oral delivery of vaccines, comprises expression vector
PT containing antigenic genomic sequence, bound to aggregated
PT protein-polycationic polymer conjugate or suspension -

```
XX
PS   Disclosure; Page 82-84; 145pp; English.
XX
CC   The invention relates to a composition comprising an expression vector
CC   bound to an aggregated protein-polycationic polymer conjugate or
CC   suspension. The expression vector contains a promoter polynucleotide
CC   sequence operatively linked to a polynucleotide sequence encoding an
CC   antigen which is a fragment of a gene or genome associated with an
CC   infectious disease, cancer and autoimmune disease such as rheumatoid
CC   arthritis, vasculitis, and multiple sclerosis, pathogenic genomes
CC   consisting of bacterium, fungus, protozoa and virus such as human
CC   immunodeficiency virus (HIV), herpes simplex virus (HSV), hepatitis C
CC   virus (HCV), influenza and respiratory syncytial virus (RSV), and
CC   optionally comprising a nucleotide sequence encoding a cytokine (or a
CC   cytokine expression vector), is useful for inducing an immune response
CC   (systemic and/or mucosal) in an organism. The cytokine expression vector
CC   contains a sequence for granulocyte macrophage-colony stimulating factor
CC   (GM-CSF) or interleukin-12 (IL-12). The polynucleotide sequences encoding
CC   the antigen and the cytokine are under transcriptional control of same or
CC   different promoter polynucleotide sequences. The expression vector, as a
CC   DNA vaccine is useful for treating a condition in an organism. The
CC   present sequence is simian-human immunodeficiency virus (SHIV) clone
CC   lA11 complete genomic DNA, related to the invention.
XX
SQ   Sequence 10274 BP; 3475 A; 1931 C; 2559 G; 2309 T; 0 other;

  Query Match             96.1%;  Score 2538.6;  DB 24;  Length 10274;
  Best Local Similarity   97.8%;  Pred. No. 0;
  Matches 2584;  Conservative    0;  Mismatches   54;  Indels    3;  Gaps    1;

Qy       1 TCTAGAAGCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTG 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6438 TCTAGAAGCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTG 6497

Qy      61 TTTTCTTAAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCC 120
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    6498 TTTTCTTAAAAAAGGCTTGGGGATATGTTATGAGCAGTCACGAAAGAGAAGAAGAACTCC 6557

Qy     121 GAAAAAGGCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGG 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6558 GAAAAAGGCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGG 6617

Qy     181 AATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATAT 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6618 AATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATAT 6677

Qy     241 GTCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCA 300
          |||||||||||||||||||||||||||||||||||||||| |||||||||| ||||||
Db    6678 GTCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGACGATTCCCCTCTTCTGTGCA 6737

Qy     301 ACCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCA 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6738 ACCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCA 6797

Qy     361 GAAGTGGCCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAG 420
          ||| |||||||||||||||||||||||||||||||| ||| | ||||||||||||||||||
Db    6798 GAATTGGCCCTTAATGTTACAGAAAGCTTTGATGCTTGGGAGAATACAGTCACAGAACAG 6857

Qy     421 GCAATAGAGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCC 480
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db    6858 GCAATAGAAGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTAACC 6917

Qy     481 CCATTATGCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAA 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6918 CCATTATGCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAA 6977

Qy     541 TCAATAACAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTC 600
          ||   | ||||||| |  || || || ||| | ||   |||| |||| || ||||| ||
Db    6978 TC---ATCAACAACAACCACAGCAGCACCAAAGGCAATGTCAGAAAAAATAAACATGATC 7034

Qy     601 AATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCAAATG 660
          |||||||||||||||||||||  || ||||||||||||||||||||||||||||||||||
Db    7035 AATGAGACTAGTTCTTGTATAGTTCATGATAATTGCACAGGCTTGGAACAAGAGCAAATG 7094

Qy     661 ATAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAATGAA 720
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7095 ATAGGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAATGAA 7154

Qy     721 ACTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGA 780
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    7155 ACTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATAGCACTGGTAATGAAAGTAGA 7214

Qy     781 TGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGG 840
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7215 TGTTACATGAATCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGG 7274

Qy     841 GATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGAC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7275 GATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGAC 7334

Qy     901 ACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7335 ACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGG 7394

Qy     961 ATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGA 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7395 ATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGA 7454

Qy    1021 ACTTATATTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTAT 1080
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    7455 ACTTATATTTACTGGCATGGTAGGGATAATAGAACTATAATTAGTTTAAATAAGTATTAT 7514

Qy    1081 AATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATG 1140
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7515 AATCTAACAATAAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATG 7574

Qy    1141 TCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGGCCAAAGCAGGCATGGTGTTGG 1200
          |||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7575 TCTCAATTGGTTTTCCACTCACAACCAATCAATGATAGGCCAAAGCAGGCATGGTGTTGG 7634

Qy    1201 TTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCC 1260
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7635 TTTAGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCC 7694

Qy    1261 AGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGAGGAGGAGAT 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7695 AGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGAGGAGGAGAT 7754

Qy    1321 CCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATGAAT 1380
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    7755 CCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTTCTCTACTGTAAAATGAAT 7814

Qy    1381 TGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGGAACAGCAT 1440
          |||||| |||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    7815 TGGTTTTTAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGGAACAACAT 7874

Qy    1441 AAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAGGC 1500
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7875 AAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAGGC 7934

Qy    1501 AAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACCAGT 1560
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    7935 AAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTATCTCCACAGTGACCAGT 7994

Qy    1561 CTCATAGCAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTGCAGAG 1620
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    7995 CTCATAGCAAACATAGATTGGACTGATGGAAACCAAACTAATATCACCATGAGTGCAGAG 8054

Qy    1621 GTGGCAGAACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTCCAATT 1680
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8055 GTGGCAGAACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTCCAATT 8114

Qy    1681 GGCTTGGCCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAAGA 1740
          |||||||||||||||| |||||| ||||||||||||||||||||||||||||||||||||
Db    8115 GGCTTGGCCCCCACAAATGTGAGGAGGTACACTACTGGTGGCACCTCAAGAAATAAAAGA 8174
```

```
Qy   1741 GGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCG 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8175 GGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCG 8234

Qy   1801 GCGTCGTTGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGCAACAG 1860
          |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   8235 GCGTCGTTGACGCTGACCGCTCAGTCCCGGACTTTATTGGCTGGGATAGTGCAGCAACAG 8294

Qy   1861 CAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCTGGGGA 1920
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8295 CAACTGCTGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCTGGGGA 8354

Qy   1921 ACAAAGAACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAG 1980
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   8355 ACAAAGAACCTCCAGACTAGGGTCACTGCCATCGAGAAATACTTAAAGGACCAGGCGCAG 8414

Qy   1981 CTGAATGCTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGCCAAAT 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8415 CTGAATGCTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGCCAAAT 8474

Qy   2041 GCAAGTCTAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGAC 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8475 GCAAGTCTAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGAC 8534

Qy   2101 TTCTTGGAAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAAC 2160
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8535 TTCTTGGAGGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAAC 8594

Qy   2161 ATGTATGAATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCT 2220
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   8595 ATGTATGAATTACAAAAGTTAAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCT 8654

Qy   2221 TCTTGGATAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGA 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8655 TCTTGGATAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGA 8714

Qy   2281 ATAGTGATCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTC 2340
          ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8715 ATAGTAATCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTC 8774

Qy   2341 TCTTCCCCACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACC 2400
          |||||||||||||||||||||| |||||||||||||||||||| ||||||||||||||||
Db   8775 TCTTCCCCACCCTCTTATTTCCAGTAGACCCATATCCAACAGGACCAGGCACTGCCAACC 8834

Qy   2401 AGAGAAGGCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAG 2460
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   8835 AGAGAAGGCAAAGAAGGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAG 8894

Qy   2461 ATAGAATATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGC 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8895 ATAGAATATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGC 8954

Qy   2521 AACTGCAGAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTC 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8955 AACTGCAGAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTC 9014

Qy   2581 TCTGCGACCCTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTACCTACAATAT 2640
          |||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||
Db   9015 TCTGCGACCCTACGAAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTACCTACAATAT 9074

Qy   2641 G 2641
          |
Db   9075 G 9075
```

```
RESULT 5
AAD38657
ID   AAD38657 standard; DNA; 6978 BP.
XX
AC   AAD38657;
XX
DT   23-SEP-2002  (first entry)
XX
DE   CMVkan/R-R-SIVgp160CTE vector DNA containing a mutated SIV env gene.
XX
KW   Gene therapy; vaccine; AIDS-related disease; immunoprophylaxis; AIDS;
KW   acquired immune deficiency syndrome; immunotherapy; immunostimulant;
KW   CMV; cytomegalovirus; simian immunodeficiency virus; SIV; chimeric;
KW   gene; mutant; CMVkan/R-R-SIVgp160CTE vector; ds.
XX
OS   Chimeric - Cytomegalovirus.
OS   Chimeric - Simian immunodeficiency virus.
OS   Chimeric - Unidentified.
XX
FH   Key             Location/Qualifiers
FT   CDS             764..3403
FT                   /*tag= a
FT                   /product= "Protein #1 encoded by CMVkan/R-R-
FT                   SIVgp160CTE vector DNA"
FT   CDS             complement (5611..6426)
FT                   /*tag= b
FT                   /product= "Protein #2 encoded by CMVkan/R-R-
FT                   SIVgp160CTE vector DNA"
XX
PN   WO200236806-A2.
XX
PD   10-MAY-2002.
XX
PF   01-NOV-2001; 2001WO-US45624.
XX
PR   01-NOV-2000; 2000US-245113P.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Pavlakis GN,  Gragerov A,  Felber BK;
XX
DR   WPI; 2002-471505/50.
DR   P-PSDB; AAE23963, AAE23964.
XX
PT   A nucleic acid construct with a sequence encoding a fusion protein
PT   comprising a destabilizing amino acid sequence attached to a
PT   heterologous amino acid, useful for gene therapy or vaccines against
PT   AIDS and AIDS-related diseases  -
XX
PS   Example 5; Fig 7; 75pp; English.
XX
CC   The present invention relates to nucleic acids (such as DNA immunisation
CC   plasmids) encoding fusion proteins containing a destabilising amino acid
CC   sequence covalently attached to a heterologous amino acid sequence, in
CC   which the immunogenicity of the heterologous amino acid sequence is
CC   increased by the presence of the destabilising amino acid sequence. The
CC   constructs are useful for in vivo or in vitro gene therapy, in the
CC   development of vaccines, in the production of diagnostic reagents,
CC   vaccines and therapies for diseases such as acquired immune deficiency
CC   syndrome (AIDS) and AIDS-related diseases. Sequences of the invention
CC   are useful in immunotherapy and immunoprophylaxis, e.g. as vaccine, or
CC   in genetic therapy after expression in mammals. The present sequence
CC   is CMV (cytomegalovirus) kan/R-R-SIV (simian immunodeficiency virus)
CC   gp160CTE vector DNA containing a mutated SIV env gene. This sequence
CC   is used in the exemplification of the invention.
XX
SQ   Sequence 6978 BP; 1973 A; 1574 C; 1685 G; 1746 T; 0 other;

  Query Match           92.2%;  Score 2434.6;  DB 24;  Length 6978;
  Best Local Similarity  98.8%;  Pred. No. 0;
  Matches 2452;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy    161 TAAGTATGGGATGTCTTGGGAATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATG 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    759 TAAGTATGGGATGTCTTGGGAATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATG 818

Qy    221 GGATCTATTGTACTCTATATGTCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGA 280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    819 GGATCTATTGTACTCTATATGTCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGA 878

Qy    281 CAATTCCCCTCTTTTGTGCAACCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTAC 340
```

```
Db    879 CAATTCCCCTCTTTTGTGCAACCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTAC 938

Qy    341 CAGATAATGGTGATTATTCAGAAGTGGCCCTTAATGTTACAGAAAGCTTTGATGCCTGGA 400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    939 CAGATAATGGTGATTATTCAGAAGTGGCCCTTAATGTTACAGAAAGCTTTGATGCCTGGA 998

Qy    401 ATAATACAGTCACAGAACAGGCAATAGAGGATGTATGGCAACTCTTTGAGACCTCAATAA 460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    999 ATAATACAGTCACAGAACAGGCAATAGAGGATGTATGGCAACTCTTTGAGACCTCAATAA 1058

Qy    461 AGCCTTGTGTAAAATTATCCCCATTATGCATTACTATGAGATGCAATAAAAGTGAGACAG 520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1059 AGCCTTGTGTAAAATTATCCCCATTATGCATTACTATGAGATGCAATAAAAGTGAGACAG 1118

Qy    521 ATAGATGGGGATTGACAAAATCAATAACAACAACAGCATCAACAACATCAACGACAGCAT 580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1119 ATAGATGGGGATTGACAAAATCAATAACAACAACAGCATCAACAACATCAACGACAGCAT 1178

Qy    581 CAGCAAAAGTAGACATGGTCAATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACAG 640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1179 CAGCAAAAGTAGACATGGTCAATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACAG 1238

Qy    641 GCTTGGAACAAGAGCAAATGATAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACA 700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1239 GCTTGGAACAAGAGCAAATGATAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACA 1298

Qy    701 AGAAAAAAGAGTACAATGAAACTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATA 760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1299 AGAAAAAAGAGTACAATGAAACTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATA 1358

Qy    761 ACACTGGTAATGAAAGTAGATGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGT 820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1359 ACACTGGTAATGAAAGTAGATGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGT 1418

Qy    821 CTTGTGACAAACATTATTGGGATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATG 880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1419 CTTGTGACAAACATTATTGGGATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATG 1478

Qy    881 CTTTGCTTAGATGTAATGACACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGG 940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1479 CTTTGCTTAGATGTAATGACACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGG 1538

Qy    941 TGGTCTCTTCATGCACAAGGATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATG 1000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1539 TGGTCTCTTCATGCACAAGGATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATG 1598

Qy   1001 GAACTAGAGCAGAAAATAGAACTTATATTTACTGGCATGGTAGGGATAATAGGACTATAA 1060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1599 GAACTAGAGCAGAAAATAGAACTTATATTTACTGGCATGGTAGGGATAATAGGACTATAA 1658

Qy   1061 TTAGTTTAAATAAGTATTATAATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAG 1120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1659 TTAGTTTAAATAAGTATTATAATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAG 1718

Qy   1121 TTTTACCAGTCACCATTATGTCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGGC 1180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1719 TTTTACCAGTCACCATTATGTCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGGC 1778

Qy   1181 CAAAGCAGGCATGGTGTTGGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGC 1240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1779 CAAAGCAGGCATGGTGTTGGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGC 1838

Qy   1241 AGACCATTGTCAAACATCCCAGGTATACTGGAACTAACAATACTGATAAAATCAATTTGA 1300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1839 AGACCATTGTCAAACATCCCAGGTATACTGGAACTAACAATACTGATAAAATCAATTTGA 1898

Qy   1301 CGGCTCCTGGAGGAGGAGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGT 1360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1899 CGGCTCCTGGAGGAGGAGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGT 1958

Qy   1361 TCCTCTACTGTAAAATGAATTGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACC 1420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   1959 TCCTCTACTGTAAAATGAATTGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACC 2018

Qy   1421 AGAAGCCAAAGGAACAGCATAAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCA 1480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2019 AGAAGCCAAAGGAACAGCATAAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCA 2078

Qy   1481 ACACTTGGCATAAAGTAGGCAAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGT 1540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2079 ACACTTGGCATAAAGTAGGCAAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGT 2138

Qy   1541 GTAACTCCACAGTGACCAGTCTCATAGCAAACATAGATTGGATTGATGGAAACCAAACTA 1600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2139 GTAACTCCACAGTGACCAGTCTCATAGCAAACATAGATTGGATTGATGGAAACCAAACTA 2198

Qy   1601 ATATCACCATGAGTGCAGAGGTGGCAGAACTGTATCGATTGGAATTGGGAGATTATAAAT 1660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2199 ATATCACCATGAGTGCAGAGGTGGCAGAACTGTATCGATTGGAATTGGGAGATTATAAAT 2258

Qy   1661 TAGTAGAGATCACTCCAATTGGCTTGGCCCCCACAGATGTGAAGAGGTACACTACTGGTG 1720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2259 TAGTAGAGATCACTCCAATTGGCTTGGCCCCCACAGATGTGAAGAGGTACACTACTGGTG 2318

Qy   1721 GCACCTCAAGAAATAAAAGAGGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGG 1780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2319 GCACCTCAAGAAATAAAAGAGGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGG 2378

Qy   1781 CAGGTTCTGCAATGGGCGCGGCGTCGTTGACGCTGACCGCTCAGTCCCGAACTTTATTGG 1840
          |||||||||||||||| || ||    |||| || || || |||||||||||||||||||
Db   2379 CAGGTTCTGCAATGGGAGCCGCCAGCCTGACCCTCACGGCACAGTCCCGAACTTTATTGG 2438

Qy   1841 CTGGGATAGTGCAGCAACAGCAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAATTGT 1900
          |||||||||| || || ||||||||||||||||||||||||||||||||||||||||||||
Db   2439 CTGGGATAGTCCAACAGCAGCAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAATTGT 2498

Qy   1901 TGCGACTGACCGTCTGGGGAACAAAGAACCTCCAGACTAGGGTCACTGCCATCGAGAAGT 1960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2499 TGCGACTGACCGTCTGGGGAACAAAGAACCTCCAGACTAGGGTCACTGCCATCGAGAAGT 2558

Qy   1961 ACTTAAAGGACCAGGCGCAGCTGAATGCTTGGGGATGTGCGTTTAGACAAGTCTGCCACA 2020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2559 ACTTAAAGGACCAGGCGCAGCTGAATGCTTGGGGATGTGCGTTTAGACAAGTCTGCCACA 2618

Qy   2021 CTACTGTACCATGGCCAAATGCAAGTCTAACACCAAAGTGGAACAATGAGACTTGGCAAG 2080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2619 CTACTGTACCATGGCCAAATGCAAGTCTAACACCAAAGTGGAACAATGAGACTTGGCAAG 2678

Qy   2081 AGTGGGAGCGAAAGGTTGACTTCTTGGAAGAAAATATAACAGCCCTCCTAGAGGAGGCAC 2140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2679 AGTGGGAGCGAAAGGTTGACTTCTTGGAAGAAAATATAACAGCCCTCCTAGAGGAGGCAC 2738

Qy   2141 AAATTCAACAAGAGAAGAACATGTATGAATTACAAAAGTTGAATAGCTGGGATGTGTTTG 2200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2739 AAATTCAACAAGAGAAGAACATGTATGAATTACAAAAGTTGAATAGCTGGGATGTGTTTG 2798

Qy   2201 GCAATTGGTTTGACCTTGCTTCTTGGATAAAGTATATACAATATGGAGTTTATATAGTTG 2260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2799 GCAATTGGTTTGACCTTGCTTCTTGGATAAAGTATATACAATATGGAGTTTATATAGTTG 2858

Qy   2261 TAGGAGTAATACTGTTAAGAATAGTGATCTATATAGTACAAATGCTAGCTAAGTTAAGGC 2320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2859 TAGGAGTAATACTGTTAAGAATAGTGATCTATATAGTACAAATGCTAGCTAAGTTAAGGC 2918

Qy   2321 AGGGGTATAGGCCAGTGTTCTCTTCCCCACCCTCTTATTTCCAGCAGACCCATATCCAAC 2380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2919 AGGGGTATAGGCCAGTGTTCTCTTCCCCACCCTCTTATTTCCAGCAGACCCATATCCAAC 2978

Qy   2381 AGGACCCGGCACTGCCAACCAGAGAAGGCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCA 2440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2979 AGGACCCGGCACTGCCAACCAGAGAAGGCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCA 3038

Qy   2441 ACAGCTCCTGGCCTTGGCAGATAGAATATATTCATTTCCTGATCCGCCAACTGATACGCC 2500
          |||||||||||||||||||||||||||||| || || || || || || || || | |
Db   3039 ACAGCTCCTGGCCTTGGCAGATAGAATATATCCACTTTCTTATTCGTCAGCTTATTAGAC 3098
```

```
Qy    2501 TCTTGACTTGGCTATTCAGCAACTGCAGAACCTTGCTATCGAGAGTATACCAGATCCTCC 2560
           |||||||||||||||||| ||||| || || ||||||||||||||||||||||||||||
Db    3099 TCTTGACTTGGCTATTCAGTAACTGTAGGACTTTGCTATCGAGAGTATACCAGATCCTCC 3158

Qy    2561 AACCAATACTCCAGAGGCTCTCTGCGACCCTACAGAGGATTCGAGAAGTCCTCAGGACTG 2620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3159 AACCAATACTCCAGAGGCTCTCTGCGACCCTACAGAGGATTCGAGAAGTCCTCAGGACTG 3218

Qy    2621 AACTGACCTACCTACAATATG 2641
           |||||||||||||||||||||
Db    3219 AACTGACCTACCTACAATATG 3239


RESULT 6
AAH22814
ID   AAH22814 standard; DNA; 7099 BP.
XX
AC   AAH22814;
XX
DT   25-SEP-2001  (first entry)
XX
DE   DNA sequence of a mutated SIV-1 env gene containing SIV vector.
XX
KW   HIV-1; gag gene; pol gene; SIV-1; env gene; mutation; immunity; AIDS;
KW   gene therapy; immune therapy; immuneprophylaxis; vaccine; retroviral;
KW   lentiviral; anti-HIV; ds.
XX
OS   Synthetic.
OS   Simian immunodeficiency virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             763..3453
FT                   /*tag= a
FT   CDS             complement (5732..6547)
FT                   /*tag= b
XX
PN   WO200146408-A2.
XX
PD   28-JUN-2001.
XX
PF   22-DEC-2000; 2000WO-US34985.
XX
PR   23-DEC-1999;   99US-0173036.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Pavlakis GN;
XX
DR   WPI; 2001-408642/43.
DR   P-PSDB; AAB85257, AAB85258.
XX
PT   Novel nucleic acid construct containing HIV-1 gag/pol and simian
PT   immunodeficiency virus gag or env genes which have been mutated to
PT   remove inhibitory/instability sequences, useful as components of
PT   anti-HIV vaccines
XX
PS   Claim 32; Fig 17; 95pp; English.
XX
CC   The invention provides nucleic acid constructs containing HIV-1 gag/pol
CC   and SIV gag or SIV env genes which have been mutated to remove or
CC   reduce inhibitory/instability sequences. Pharmaceutical compositions
CC   comprising the nucleic acid construct and a carrier is useful for
CC   inducing antibodies or cytotoxic T lymphocytes in a mammal. Vaccine
CC   compositions are useful for inducing immunity against HIV-1 infection in
CC   a human. The constructs are useful for producing infectious viral
CC   particles. The vectors, host cells and vector systems comprising the
CC   constructs are useful in gene therapy, immune therapy, immuneprophylaxis,
CC   vaccine or in genetic therapy after expression. The complete removal of
CC   the inhibitory/instability sequence elements from the lentiviral genome
CC   would result in constitutive expression. This could prevent the virus
CC   from establishing a latent infection and escaping immune system
CC   surveillance. The constructs are also useful for in vivo or in vitro gene
CC   replacement. Mutated genes can be expressed in the native host cell or
CC   organism or in a different cell or organism. The mutated genes can be
CC   introduced into a vector such as a plasmid, cosmid, minichromosome or
CC   virus and inserted into a host cell or organism. Host cells containing
CC   these mutated genes can be used to express large amounts of the protein
CC   useful in enzyme preparations, pharmaceuticals, diagnostic reagents,
CC   vaccines and therapeutics. The constructs and/or proteins made by using
CC   constructs encoding the mutated gag, env, and pol genes could be used
CC   in the production of diagnostic reagents, and therapies for AIDS and
CC   AIDS related diseases. Further, the gag constructs could be used in
CC   transdominant inhibition of HIV expression after the introduction into
CC   humans. The vectors would be introduced into humans, resulting in the
CC   inhibition and of immunostimulation by the produced gag protein. In
CC   addition, the gag constructs could be used in the generation of new
CC   retroviral vectors based on the expression of lentiviral gag proteins.
CC   The present sequence represents the DNA sequence of the SIV envelope
CC   vector CMVkan/R-R-SIVenvCTE containing a mutated SIV-1 env gene.
XX
SQ   Sequence 7099 BP; 2003 A; 1597 C; 1730 G; 1769 T; 0 other;

  Query Match             92.2%;  Score 2434.6;  DB 22;  Length 7099;
  Best Local Similarity   98.8%;  Pred. No.[1] 0;
  Matches 2452;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy     161 TAAGTATGGGATGTCTTGGGAATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATG 220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     758 TAAGTATGGGATGTCTTGGGAATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATG 817

Qy     221 GGATCTATTGTACTCTATATGTCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGA 280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     818 GGATCTATTGTACTCTATATGTCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGA 877

Qy     281 CAATTCCCCTCTTTTGTGCAACCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTAC 340
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     878 CAATTCCCCTCTTTTGTGCAACCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTAC 937

Qy     341 CAGATAATGGTGATTATTCAGAAGTGGCCCTTAATGTTACAGAAAGCTTTGATGCCTGGA 400
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     938 CAGATAATGGTGATTATTCAGAAGTGGCCCTTAATGTTACAGAAAGCTTTGATGCCTGGA 997

Qy     401 ATAATACAGTCACAGAACAGGCAATAGAGGATGTATGGCAACTCTTTGAGACCTCAATAA 460
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     998 ATAATACAGTCACAGAACAGGCAATAGAGGATGTATGGCAACTCTTTGAGACCTCAATAA 1057

Qy     461 AGCCTTGTGTAAAATTATCCCCATTATGCATTACTATGAGATGCAATAAAAGTGAGACAG 520
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1058 AGCCTTGTGTAAAATTATCCCCATTATGCATTACTATGAGATGCAATAAAAGTGAGACAG 1117

Qy     521 ATAGATGGGGATTGACAAAATCAATAACAACAACAGCATCAACAACATCAACGACAGCAT 580
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1118 ATAGATGGGGATTGACAAAATCAATAACAACAACAGCATCAACAACATCAACGACAGCAT 1177

Qy     581 CAGCAAAAGTAGACATGGTCAATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACAG 640
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1178 CAGCAAAAGTAGACATGGTCAATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACAG 1237

Qy     641 GCTTGGAACAAGAGCAAATGATAAGCTGTAAATTCAACATGACAGGGTTAAAAACAGACA 700
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1238 GCTTGGAACAAGAGCAAATGATAAGCTGTAAATTCAACATGACAGGGTTAAAAACAGACA 1297

Qy     701 AGAAAAAAGAGTACAATGAAACTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATA 760
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1298 AGAAAAAAGAGTACAATGAAACTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATA 1357

Qy     761 ACACTGGTAATGAAAGTAGATGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGT 820
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1358 ACACTGGTAATGAAAGTAGATGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGT 1417

Qy     821 CTTGTGACAAACATTATTGGGATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATG 880
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1418 CTTGTGACAAACATTATTGGGATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATG 1477

Qy     881 CTTTGCTTAGATGTAATGACACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGG 940
```

```
Db   1478 CTTTGCTTAGATGTAATGACACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGG 1537

Qy    941 TGGTCTCTTCATGCACAAGGATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATG 1000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1538 TGGTCTCTTCATGCACAAGGATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATG 1597

Qy   1001 GAACTAGAGCAGAAAATAGAACTTATATTTACTGGCATGGTAGGGATAATAGGACTATAA 1060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1598 GAACTAGAGCAGAAAATAGAACTTATATTTACTGGCATGGTAGGGATAATAGGACTATAA 1657

Qy   1061 TTAGTTTAAATAAGTATTATAATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAG 1120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1658 TTAGTTTAAATAAGTATTATAATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAG 1717

Qy   1121 TTTTACCAGTCACCATTATGTCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGGC 1180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1718 TTTTACCAGTCACCATTATGTCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGGC 1777

Qy   1181 CAAAGCAGGCATGGTGTTGGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGC 1240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1778 CAAAGCAGGCATGGTGTTGGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGC 1837

Qy   1241 AGACCATTGTCAAACATCCCAGGTATACTGGAACTAACAATACTGATAAAATCAATTTGA 1300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1838 AGACCATTGTCAAACATCCCAGGTATACTGGAACTAACAATACTGATAAAATCAATTTGA 1897

Qy   1301 CGGCTCCTGGAGGAGGAGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGT 1360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1898 CGGCTCCTGGAGGAGGAGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGT 1957

Qy   1361 TCCTCTACTGTAAAATGAATTGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACC 1420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1958 TCCTCTACTGTAAAATGAATTGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACC 2017

Qy   1421 AGAAGCCAAAGGAACAGCATAAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCA 1480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2018 AGAAGCCAAAGGAACAGCATAAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCA 2077

Qy   1481 ACACTTGGCATAAAGTAGGCAAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGT 1540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2078 ACACTTGGCATAAAGTAGGCAAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGT 2137

Qy   1541 GTAACTCCACAGTGACCAGTCTCATAGCAAACATAGATTGGATTGATGGAAACCAAACTA 1600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2138 GTAACTCCACAGTGACCAGTCTCATAGCAAACATAGATTGGATTGATGGAAACCAAACTA 2197

Qy   1601 ATATCACCATGAGTGCAGAGGTGGCAGAACTGTATCGATTGGAATTGGGAGATTATAAAT 1660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2198 ATATCACCATGAGTGCAGAGGTGGCAGAACTGTATCGATTGGAATTGGGAGATTATAAAT 2257

Qy   1661 TAGTAGAGATCACTCCAATTGGCTTGGCCCCCACAGATGTGAAGAGGTACACTACTGGTG 1720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2258 TAGTAGAGATCACTCCAATTGGCTTGGCCCCCACAGATGTGAAGAGGTACACTACTGGTG 2317

Qy   1721 GCACCTCAAGAAATAAAAGAGGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGG 1780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2318 GCACCTCAAGAAATAAAAGAGGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGG 2377

Qy   1781 CAGGTTCTGCAATGGGCGCGGCGTCGTTGACGCTGACCGCTCAGTCCCGAACTTTATTGG 1840
          |||||||||||||||| || ||     |||| || || || ||||||||||||||||||||
Db   2378 CAGGTTCTGCAATGGGAGCCGCCAGCCTGACCCTCACGGCACAGTCCCGAACTTTATTGG 2437

Qy   1841 CTGGGATAGTGCAGCAACAGCAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAATTGT 1900
          |||||||||| || || ||||||||||||||||||||||||||||||||||||||||||||
Db   2438 CTGGGATAGTCCAACAGCAGCAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAATTGT 2497

Qy   1901 TGCGACTGACCGTCTGGGGAACAAAGAACCTCCAGACTAGGGTCACTGCCATCGAGAAGT 1960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2498 TGCGACTGACCGTCTGGGGAACAAAGAACCTCCAGACTAGGGTCACTGCCATCGAGAAGT 2557

Qy   1961 ACTTAAAGGACCAGGCGCAGCTGAATGCTTGGGGATGTGCGTTTAGACAAGTCTGCCACA 2020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   2558 ACTTAAAGGACCAGGCGCAGCTGAATGCTTGGGGATGTGCGTTTAGACAAGTCTGCCACA 2617

Qy   2021 CTACTGTACCATGGCCAAATGCAAGTCTAACACCAAAGTGGAACAATGAGACTTGGCAAG 2080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2618 CTACTGTACCATGGCCAAATGCAAGTCTAACACCAAAGTGGAACAATGAGACTTGGCAAG 2677

Qy   2081 AGTGGGAGCGAAAGGTTGACTTCTTGGAAGAAAATATAACAGCCCTCCTAGAGGAGGCAC 2140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2678 AGTGGGAGCGAAAGGTTGACTTCTTGGAAGAAAATATAACAGCCCTCCTAGAGGAGGCAC 2737

Qy   2141 AAATTCAACAAGAGAAGAACATGTATGAATTACAAAAGTTGAATAGCTGGGATGTGTTTG 2200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2738 AAATTCAACAAGAGAAGAACATGTATGAATTACAAAAGTTGAATAGCTGGGATGTGTTTG 2797

Qy   2201 GCAATTGGTTTGACCTTGCTTCTTGGATAAAGTATATACAATATGGAGTTTATATAGTTG 2260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2798 GCAATTGGTTTGACCTTGCTTCTTGGATAAAGTATATACAATATGGAGTTTATATAGTTG 2857

Qy   2261 TAGGAGTAATACTGTTAAGAATAGTGATCTATATAGTACAAATGCTAGCTAAGTTAAGGC 2320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2858 TAGGAGTAATACTGTTAAGAATAGTGATCTATATAGTACAAATGCTAGCTAAGTTAAGGC 2917

Qy   2321 AGGGGTATAGGCCAGTGTTCTCTTCCCCACCCTCTTATTTCCAGCAGACCCATATCCAAC 2380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2918 AGGGGTATAGGCCAGTGTTCTCTTCCCCACCCTCTTATTTCCAGCAGACCCATATCCAAC 2977

Qy   2381 AGGACCCGGCACTGCCAACCAGAGAAGGCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCA 2440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2978 AGGACCCGGCACTGCCAACCAGAGAAGGCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCA 3037

Qy   2441 ACAGCTCCTGGCCTTGGCAGATAGAATATATTCATTTCCTGATCCGCCAACTGATACGCC 2500
          |||||||||||||||||||||||||||||| || || || || || || || || | |
Db   3038 ACAGCTCCTGGCCTTGGCAGATAGAATATATCCACTTTCTTATTCGTCAGCTTATTAGAC 3097

Qy   2501 TCTTGACTTGGCTATTCAGCAACTGCAGAACCTTGCTATCGAGAGTATACCAGATCCTCC 2560
          |||||||||||||||||| ||||| || || ||||||||||||||||||||||||||||
Db   3098 TCTTGACTTGGCTATTCAGTAACTGTAGGACTTTGCTATCGAGAGTATACCAGATCCTCC 3157

Qy   2561 AACCAATACTCCAGAGGCTCTCTGCGACCCTACAGAGGATTCGAGAAGTCCTCAGGACTG 2620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3158 AACCAATACTCCAGAGGCTCTCTGCGACCCTACAGAGGATTCGAGAAGTCCTCAGGACTG 3217

Qy   2621 AACTGACCTACCTACAATATG 2641
          |||||||||||||||||||||
Db   3218 AACTGACCTACCTACAATATG 3238
```

```
RESULT 7
AAN80860
ID   AAN80860 standard; cDNA; 9601 BP.
XX
AC   AAN80860;
XX
DT   15-NOV-1990  (first entry)
XX
DE   Sequence corresp. to the genome of SIV isolate STLV-III Mac 142-83
DE   in clone SIV-1.
XX
KW   pSIV-1.1; pSIV-1.2; STLV-III Mac 142-83; ss.
XX
OS   Simian immunodeficiency virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             551..2068
FT                   /*tag=  a
FT                   /gene="gagd"
FT   CDS             1726..4893
FT                   /*tag=  b
FT                   /gene="pol"
FT   CDS             4826..5467
FT                   /*tag=  c
FT                   /gene="Q"
FT   CDS             5298..5633
```

```
FT                   /*tag= d
FT                   /gene="X"
FT   CDS             5637..5939
FT                   /*tag= e
FT                   /gene="R"
FT   CDS             5788..6084
FT                   /*tag= f
FT                   /gene="tat1"
FT   CDS             6014..6130
FT                   /*tag= g
FT                   /gene="art1"
FT   CDS             6090..8297
FT                   /*tag= h
FT                   /gene="env"
FT   CDS             8298..8732
FT                   /*tag= i
FT                   /gene="env ctd."
FT   CDS             8296..8391
FT                   /*tag= j
FT                   /gene="tat2"
FT   CDS             8294..8548
FT                   /*tag= k
FT                   /gene="art2"
FT   CDS             8569..9354
FT                   /*tag= l
FT                   /gene="F"
XX
PN   WO8805440-A.
XX
PD   28-JUL-1988.
XX
PF   15-JAN-1988;   88WO-FR00025.
XX
PR   15-APR-1987;   87FR-0005398.
XX
PA   (INSP ) INST PASTEUR.
PA   (ALIZ/) ALIZON M.
XX
PI   Alizon M, Montagnier L, Guetard D, Clavfl F, Sonigo P, Guyader M;
PI   Tiollais P, Chakrabarti L, Desrosiers R;
XX
DR   WPI; 1988-220290/31.
DR   P-PSDB; AAP80802, AAP81783, AAP81784, AAP81785, AAP81691, AAP81756,
DR   AAP81757, AAP81758, AAP81759, AAP81760, AAP81761, AAP81762.
XX
PT   New peptide(s) with immunological properties of HIV-2 envelope protein -
PT   having the structure of simian immune deficiency virus proteins,
PT   useful in diagnosis and of vaccine components
XX
PS   Claim 23; Fig 1B; 86pp; French.
XX
CC   AAN80860 is contained in lambda SIV-1 which was deposited on 15/4/87 at
CC   the CNCM under numbers I-658 (pSIV-1.1) and I-659 (pSIV-1.2). It is the
CC   cDNA to SIV-1 genomic RNA. Recombinant DNA contg. all or part of the
CC   cDNA from this sequence inserted into a vector are claimed. Labelled
CC   fragments of cDNA can be used as hybridisation probes. New peptides which
CC   have immunological properties in common with those of the peptide
CC   skeleton of the envelope protein of HIV-2 and also have a peptide
CC   structure in common with that of SIV-1 glycoprotein are claimed.
CC   Antigenic and immunogenic conjugates contg. the peptides and a kit to
CC   detect HIV-2 in biological fluids are new. The peptides are useful for in
CC   vitro diagnosis of HIV-2 infection and some of them can be used as
CC   components of immunogens and vaccines against HIV. Antibodies raised
CC   against them can be used for treatment of AIDS.
XX
SQ   Sequence 9601 BP; 3277 A; 1809 C; 2378 G; 2136 T; 1 other;

  Query Match             90.6%;  Score 2392.8;  DB 9;  Length 9601;
  Best Local Similarity   94.8%;  Pred. No. 0;
  Matches 2511;  Conservative    0;  Mismatches  127;  Indels   12;  Gaps    3;

Qy      1 TCTAGAAGCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTG 60
          ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
```

```
Db   5926 TCTAGAAGCATGCTATAACACATGCTATTGCAAAAAGTGTTGCTACCATTGCCAGTTTTG 5985

Qy     61 TTTTCTTAAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCC 120
          |||||||||||| |||||||||||| |||||||| | ||||  | ||||||||||||||||
Db   5986 TTTTCTTAAAAAGGGCTTGGGGATAAGTTATGAGAAGTCACACAGGAGAAGAAGAACTCC 6045

Qy    121 GAAAAAGGCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGG 180
          ||| |||||||||||||||||||||||||||||||||||||| |||||||||||| ||||||||
Db   6046 GAAGAAGGCTAAGGCTAATACATCTTCTGCATCAAACGAGTAAGTATGGGTTGTCTTGGA 6105

Qy    181 AATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATAT 240
          |||||||||||||||||||||    | |||||||||| |||| |||||| || ||||
Db   6106 AATCAGCTGCTTATCGCCATCT--GCTCTAAGTGTCTAT-GGATTATTTGTATTCAATAT 6162

Qy    241 GTCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCA 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   6163 GTCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTCTGTGCA 6222

Qy    301 ACCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCA 360
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   6223 ACCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGATGATTATTCA 6282

Qy    361 GAAGTGGCCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAG 420
          ||| |||||||||||||||||||||||||||||| ||| | ||||||||||||||||||
Db   6283 GAATTGGCCCTTAATGTTACAGAAAGCTTTGATGCTTGGGAGAATACAGTCACAGAACAG 6342

Qy    421 GCAATAGAGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCC 480
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6343 GCAATAGAGGACGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCC 6402

Qy    481 CCATTATGCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAA 540
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   6403 CCATTATGCATTACTATGAGATGCAATAAAAGTGAGACAGATAAATGGGGATTGACAAAA 6462

Qy    541 TCAATAACAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGT---------A 591
          |||  |||||||||||||||||||| |||| ||| || ||  |  |||         |
Db   6463 TCATCAACAACAACAGCATCAACAACAACAACAACAGCAAAATCAGTAGAGACAAGA 6522

Qy    592 GACATGGTCAATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAA 651
          ||||| |||||||||||||||| ||||| ||| || || ||||||||||||||||||||||
Db   6523 GACATAGTCAATGAGACTAGTCCTTGTGTAGTTCATGATAATTGCACAGGCTTGGAACAA 6582

Qy    652 GAGCAAATGATAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAG 711
          |||| |||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   6583 GAGCCAATGATAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAGGAG 6642

Qy    712 TACAATGAAACTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAAT 771
          |||||||||||||||||||||||||| |||| |||||||||||||||| ||||||||||
Db   6643 TACAATGAAACTTGGTACTCTGCAGATCTGGTTTGTGAACAAGGGAATAGCACTGGTAAT 6702

Qy    772 GAAAGTAGATGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAA 831
          |||||||||||||||||||| |||||||| |||||||||||||||||| ||||||||||
Db   6703 GAAAGTAGATGTTACATGAATCACTGTAATACTTCTGTTATCCAAGAGTGTTGTGACAAA 6762

Qy    832 CATTATTGGGATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGA 891
          |||||||||||||||||||| ||| ||||||||||||||||||||||||||||||||||||
Db   6763 GATTATTGGGATGCTATTAGATGTAGATATTGTGCACCTCCAGGTTATGCTTTGCTTAGA 6822

Qy    892 TGTAATGACACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCA 951
          |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db   6823 TGTAATGACACAAATTATTCAGGCTTTATGCCTAACTGTTCTAAGGTAGTGGTCTCTTCA 6882

Qy    952 TGCACAAGGATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCA 1011
          |||||||||||||||||||||||||||||||||||| | ||||||||||||||||||||
Db   6883 TGCACAAGGATGATGGAGACACAGACTTCTACTTGGTTTCGGTTTAATGGAACTAGAGCA 6942

Qy   1012 GAAAATAGAACTTATATTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAAT 1071
          |||||||||| |||||||||||||||||||| |||||||||||||||||||||||||||||
Db   6943 GAAAATAGAACCTATATTTACTGGCATGGTAGAGATAATAGGACTATAATTAGTCTAAAT 7002

Qy   1072 AAGTATTATAATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTC 1131
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7003 AAGCATTATAATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTC 7062
```

```
Qy   1132 ACCATTATGTCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGGCCAAAGCAGGCA 1191
          |||||||||||||| |||||||||||||||||||||| ||||||| ||||||||||||||||
Db   7063 ACCATTATGTCTGCATTGGTTTTCCACTCACAACCAGTCAATGAGAGGCCAAAGCAGGCA 7122

Qy   1192 TGGTGTTGGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTC 1251
          |||||| |||||||||||||| |||||||| |||||||||||||||||||||||||||||
Db   7123 TGGTGTAGGTTTGGAGGAAATTGGAAGGAGGCAATAAAAGAGGTGAAGCAGACCATTGTC 7182

Qy   1252 AAACATCCCAGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGA 1311
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db   7183 AAACATCCCAGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTAGA 7242

Qy   1312 GGAGGAGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGT 1371
          |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db   7243 GGAGGAGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTTCTCTACTGT 7302

Qy   1372 AAAATGAATTGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAG 1431
          |||||||||||||||||||||||||||||||||||||||| |  | ||| ||||||||||||
Db   7303 AAAATGAATTGGTTTCTAAATTGGGTAGAAGATAGGAGTCTAACTACCCAGAAGCCAAAG 7362

Qy   1432 GAACAGCATAAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCAT 1491
          |||| |||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   7363 GAACGGCATAAAAGGAATTACGTACCATGTCATATTAGACAAATAATCAACACTTGGCAT 7422

Qy   1492 AAAGTAGGCAAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACA 1551
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7423 AAAGTAGGCAAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACA 7482

Qy   1552 GTGACCAGTCTCATAGCAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATG 1611
          |||||||||||||||||||||| |||||| |||||||||||||||| ||||||||||
Db   7483 GTGACCAGTCTCATAGCAAACATAAATTGGACTGATGGAAACCAAACTAGTATCACCATG 7542

Qy   1612 AGTGCAGAGGTGGCAGAACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATC 1671
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   7543 AGTGCAGAGGTGGCAGAACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAAATC 7602

Qy   1672 ACTCCAATTGGCTTGGCCCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGA 1731
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   7603 ACTCCAATTGGCTTGGCCCCCACAAATGTGAAGAGGTACACTACTGGTGGCACCTCAAGA 7662

Qy   1732 AATAAAAGAGGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCA 1791
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7663 AATAAAAGAGGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCA 7722

Qy   1792 ATGGGCGCGGCGTCGTTGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTG 1851
          |||||||||||||||||||  |||||||||||||||| ||||||||||||||||||||||
Db   7723 ATGGGCGCGGCGTCGTTGACCGTGACCGCTCAGTCCCGGACTTTATTGGCTGGGATAGTG 7782

Qy   1852 CAGCAACAGCAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACC 1911
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7783 CAGCAACAGCAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACC 7842

Qy   1912 GTCTGGGGAACAAAGAACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGAC 1971
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   7843 GTCTGGGGAACAAAGAACCTCCAGACTAGGGTCTCTGCCATCGAGAAGTACTTAAAGGAC 7902

Qy   1972 CAGGCGCAGCTGAATGCTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCA 2031
          |||||||||| |||||||||||||||||||||||||||||||| ||||||||||||||||
Db   7903 CAGGCGCAGCTAAATGCTTGGGGATGTGCGTTTAGACAAGTCTGTCACACTACTGTACCA 7962

Qy   2032 TGGCCAAATGCAAGTCTAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGA 2091
          |||||||||||||||||||||| | |||||||||||||||||||||||||||||||||||
Db   7963 TGGCCAAATGCAAGTCTAACACCAGATTGGAACAATGAGACTTGGCAAGAGTGGGAGCGG 8022

Qy   2092 AAGGTTGACTTCTTGGAAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAA 2151
          |||||||||||||||| | ||||||||| |||||||||| ||||||||||||||||||||
Db   8023 AAGGTTGACTTCTTGGAGGCAAATATAACGGCCCTCCTAGAAGAGGCACAAATTCAACAA 8082

Qy   2152 GAGAAGAACATGTATGAATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTT 2211
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8083 GAGAAGAACATGTATGAATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTT 8142

Qy   2212 GACCTTGCTTCTTGGATAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATA 2271
          |||||| |||||||||||||||||||||||||||||| |||||||| ||||||||||||||
Db   8143 GACCTTACTTCTTGGATAAAGTATATACAATATGGAATTTATATAATTGTAGGAGTAATA 8202

Qy   2272 CTGTTAAGAATAGTGATCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGG 2331
          |||||||||||||||||||||||||||||||||||||||||||| |||||| ||||||||||||
Db   8203 CTGTTAAGAATAGTGATCTATATAGTACAAATGCTAGCTAGGTTAAGACAGGGGTATAGG 8262

Qy   2332 CCAGTGTTCTCTTCCCCACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCA 2391
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||| |||||
Db   8263 CCAGTGTTCTCTTCCCCACCCTCTTATTTCCAGTAGACCCATACCCAACAGGATCCGGCT 8322

Qy   2392 CTGCCAACCAGAGAAGGCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGG 2451
          ||||||||| |||||||||| || |||||||||||   ||||||||||||||||||||||
Db   8323 CTGCCAACCAAAGAAGGCAAAAAAGGAGACGGTGGAGGCAGCGGTGGCAACAGCTCCTGG 8382

Qy   2452 CCTTGGCAGATAGAATATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGG 2511
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8383 CCTTGGCAGATAGAATATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGG 8442

Qy   2512 CTATTCAGCAACTGCAGAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTC 2571
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||| ||
Db   8443 CTATTCAGCAACTGCAGAACCTTGCTATCGAGAGCATACCAGATCCTCCAACCAATATTC 8502

Qy   2572 CAGAGGCTCTCTGCGACCCTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTAC 2631
          |||||||||||||||||          | ||||||||||||  ||||||||||||
Db   8503 CAGAGGCTCTCTGCGACCTACGGAGAATTCGGAGAAGTCCTCAGGCTTGAACTGACCTAC 8562

Qy   2632 CTACAATATG 2641
          ||||||||||
Db   8563 CTACAATATG 8572
```

RESULT 8
AAT09009
ID AAT09009 standard; DNA; 2658 BP.
XX
AC AAT09009;
XX
DT 05-JUL-1996 (first entry)
XX
DE SIV subgenomic fragment contg. rev and envelope (gp160) sequences.
XX
KW adeno-associated virus 2; AAV2; inverted terminal repeat; viral vector;
KW DNA delivery; immunodeficiency virus protein; immunity; human; simian;
KW neurological disorder; therapy; ss.
XX
OS Simian immunodeficiency virus.
XX
PN WO9534670-A2.
XX
PD 21-DEC-1995.
XX
PF 06-JUN-1995; 95WO-US07178.
XX
PR 06-JUN-1994; 94US-0254358.
XX
PA (CHIL-) CHILDRENS HOSPITAL INC.
XX
PI Johnson PR;
XX
DR WPI; 1996-049697/05.
XX
PT Recombinant adeno-associated virus genome contg. protein encoding
PT DNA - flanked by inverted terminal repeats, for use in vaccines or
PT for treatment of neuro-degenerative disease
XX
PS Example 1; Page 30-31; 44pp; English.
XX
CC The present sequence is that of a PCR-generated, 2.7 kb Simian immuno-
CC deficiency virus (SIV) subgenomic fragment contg. the rev and envelope
CC gp160) sequences. This sequence was cloned into an unique BamHI site
CC in the vector pAAV/CMV, comprising the Adeno-associated virus (AAV)

```
CC   inverted terminal repeats (ITRs) and a CMV expression cassette contg. a
CC   portion of the CMV immediate early promoter, a 640 bp intron and a
CC   polyadenylation signal sequence which were derived from the rabbit
CC   beta-globin gene.  The unique BamHI site was between the intron and
CC   polyA sequences. The resulting recombinant AAV genome vector,
CC   pAAV/CMV/SIVrev -gp160, is 8.53 kb in length.  In transient transfection
CC   assays of human 293 cells, this vector resulted in high levels of
CC   expression of the SIV gp160 protein as determined by
CC   radioimmunoprecipitation assays using polyclonal sera from monkeys
CC   infected with SIV.  The SIV rev/envelope sequences can be replaced by,
CC   e.g. HIV-1 (human immunodeficiency virus) rev/envelope sequences (see
CC   AAT09010), nerve growth factor, ciliary neuro- trophic factor and glial
CC   cell derived neurotrophic factor, etc..
XX
SQ   Sequence 2658 BP; 864 A; 480 C; 656 G; 658 T; 0 other;

  Query Match             66.1%;  Score 1745.4;  DB 17;  Length 2658;
  Best Local Similarity   82.2%;  Pred. No. 0;
  Matches 2052;  Conservative    0;  Mismatches  421;  Indels   24;  Gaps    3;

Qy    166 ATGGGATGTCTTGGGAATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATC 225
          ||||||||||||||||||||||||||||||||  |||||||  |||||||  |  | |||
Db      1 ATGGGATGTCTTGGGAATCAGCTGCTTATCGCGCTCTTGCTAGTAAGTGTTTTAGAGATT 60

Qy    226 TATTGTACTCTATATGTCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATT 285
          | ||||  || ||||||  |||||  || ||||||||||||||  ||||  ||||||||||||
Db     61 TGTTGTGTTCAATATGTAACAGTATTCTATGGTGTACCAGCATGGAAGAATGCGACAATT 120

Qy    286 CCCCTCTTTTGTGCAACCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGAT 345
          ||||||||  ||||||||||||||||||||  ||||||||||||||  || |||  | ||||||
Db    121 CCCCTCTTCTGTGCAACCAAGAATAGGGACACTTGGGGAACAACACAATGCTTGCCAGAT 180

Qy    346 AATGGTGATTATTCAGAAGTGGCCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAAT 405
          ||||  ||||||  ||||||  ||||   | |||||  |||||    ||||||||  ||| |||||
Db    181 AATGATGATTACTCAGAATTGGCAATCAATGTCACAGAGGCTTTTGATGCTTGGGATAAT 240

Qy    406 ACAGTCACAGAACAGGCAATAGAGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCT 465
          |||||||||||||  ||||||||||||  |||  | ||||||||  ||  ||  ||  |||||
Db    241 ACAGTCACAGAACAAGCAATAGAGGATGTGTGGAACCTCTTTGAAACATCCATTAAGCCC 300

Qy    466 TGTGTAAAATTATCCCCATTATGCATTACTATGAGATGCAATAAAAGTGAGACAGATAGA 525
          |||||||||  |  |||||  ||||  ||  | ||||||||  |||||||  ||||||||||||
Db    301 TGTGTAAAACTCACCCCACTATGTATAGCAATGAGATGTAATAAAACTGAGACAGATAGG 360

Qy    526 TGGGGATTGACAAAATCA------ATAACAACAACAGCATCAACAACATCAACGACAGCA 579
          |||||  ||||||   |        | ||||||||||||||  |||||||  ||  | |||  ||
Db    361 TGGGGTTTGACAGGAAACGCAGGGACAACAACAACAGCAATAACAACAACAGCAACACCA 420

Qy    580 TCAGCAAAAGTAGACATGGTCAATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACA 639
          |  |  ||  |  |   |  |||||  |  ||  |  |  |||     |  |||  |||  ||
Db    421 AGTGTAGCAGAAAATGTTATAAATGAAAGTAATCCGGGCATAAAAAATAATAGTTGTGCA 480

Qy    640 GGCTTGGAACAAGAGCAAATGATAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGAC 699
          ||||||||||  ||||  ||||||  | ||||||||  ||||||||||||||||  || |||
Db    481 GGCTTGGAACAGGAGCCCATGATAGGTTGTAAATTTAACATGACAGGGTTAAATAGGGAC 540

Qy    700 AAGAAAAAAGAGTACAATGAAACTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAAT 759
          || ||  |||||  ||  ||||||||  |||||  ||    ||||||   | |||||  ||   |
Db    541 AAAAAGAAAGAATATAATGAAACATGGTATTCAAGAGATTTAATCTGTGAGCA---GTCA 597

Qy    760 AACACTGGTAATGAAAGTAGATGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAG 819
          |  ||   | |||  ||||  ||||||||  |  ||  |||||||    ||||||  |||||
Db    598 GCGAATGAAAGTGAGAGTAAATGTTACATGCATCATTGTAACACCAGTGTTATTCAAGAA 657

Qy    820 TCTTGTGACAAACATTATTGGGATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTAT 879
          ||  ||||||||  ||||||||||||||||||||||||||  ||  ||||||||  |||||||||
Db    658 TCCTGTGACAAGCATTATTGGGATGCTATTAGATTTAGATACTGTGCACCGCCAGGTTAT 717

Qy    880 GCTTTGCTTAGATGTAATGACACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTG 939
          ||||||||||  ||||||||  |||||||||  |||||||    |||||  ||||||||||
Db    718 GCTTTGCTTAGGTGTAATGATTCAAATTATTTAGGCTTTGCTCCTAACTGTTCTAAGGTA 777

Qy    940 GTGGTCTCTTCATGCACAAGGATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAAT 999
          |||||  ||||||||||||||  ||||||||||  || ||  ||||||||||||||||  |||
Db    778 GTGGTTTCTTCATGCACAAGAATGATGGAGACGCAAACCTCTACTTGGTTTGGCTTCAAT 837

Qy   1000 GGAACTAGAGCAGAAAATAGAACTTATATTTACTGGCATGGTAGGGATAATAGGACTATA 1059
          || |||||  ||||||||||||||  || |||||  ||||||||  |    ||||||  || |||
Db    838 GGTACTAGGGCAGAAAATAGAACATACATTTATTGGCATGGCAAAAGTAATAGAACCATA 897

Qy   1060 ATTAGTTTAAATAAGTATTATAATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACA 1119
          |||||  ||  ||||||||||||||||||||||||||||||  ||||||||||||  ||||||||||
Db    898 ATTAGCTTGAATAAGTATTATAATCTAACAATGAGATGTAGAAGACCAGAAAATAAGACA 957

Qy   1120 GTTTTACCAGTCACCATTATGTCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGG 1179
          ||||||||||||||||||||||||  ||  |||||  |||||  ||  ||  ||  ||||  ||
Db    958 GTTTTACCAGTCACCATTATGTCAGGGTTGGTCTTCCATTCGCAGCCCATAAATGAGAGA 1017

Qy   1180 CCAAAGCAGGCATGGTGTTGGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAG 1239
          |||||  |||||  |||||  |||||||  |||||  |||||  |  ||  ||  |  || ||||||
Db    1018 CCAAAACAGGCCTGGTGCTGGTTTGAAGGAAGCTGGAAAAAGGCCATCCAGGAAGTGAAG 1077

Qy   1240 CAGACCATTGTCAAACATCCCAGGTATACTGGAACTAACAATACTGATAAAATCAATTTG 1299
          | |||  | ||||||||||||||||||||  ||||||||  |||||  |||||  |||  |
Db    1078 GAAACCTTGGTCAAACATCCCAGGTATACGGGAACTAATGATACTAGGAAAATTAATCTA 1137

Qy   1300 ACGGCTCCTGGAGGAGGAGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAG 1359
          ||  |||||  | ||||||||||||  |||||  ||  ||  ||||||||||||||  |||||||
Db    1138 ACAGCTCCAGCAGGAGGAGATCCAGAAGTCACTTTTATGTGGACAAATTGTCGAGGAGAA 1197

Qy   1360 TTCCTCTACTGTAAAATGAATTGGTTTCTAAATTGGGTAGAAGATAGGAATACA------ 1413
          |||  |  ||  ||  ||||||||||||||||||  ||||||||||  ||  ||  |    |
Db    1198 TTCTTATATTGCAAAATGAATTGGTTTCTTAATTGGGTAGAGGACAGAGACCAAAAGGGT 1257

Qy   1414 ---------GCTAACCAGAAGCCAAAGGAACAGCATAAAAGGAATTACGTGCCATGTCAT 1464
                    | ||  ||    ||  ||  ||  ||  ||  |  |||||  ||||||||||||
Db    1258 GGCAGATGGAAACAACAAAATAGGAAAGAGCAACAGAAGAAAAATTATGTGCCATGTCAT 1317

Qy   1465 ATTAGACAAATAATCAACACTTGGCATAAAGTAGGCAAAAATGTTTATTTGCCTCCAAGA 1524
          ||||||||||||||||||  |||||  ||||||||||||||||  ||||||||||  ||
Db    1318 ATTAGACAAATAATCAACACGTGGCACAAAGTAGGCAAAAATGTATATTTGCCTCCTAGG 1377

Qy   1525 GAGGGAGACCTCACGTGTAACTCCACAGTGACCAGTCTCATAGCAAACATAGATTGGATT 1584
          ||  ||||||||  ||  ||  |||||  ||  ||  ||||||||||||  | ||||||||||
Db    1378 GAAGGAGACCTGACATGCAATTCCACTGTAACTAGTCTCATAGCAGAGATAGATTGGATC 1437

Qy   1585 GATGGAAACCAAACTAATATCACCATGAGTGCAGAGGTGGCAGAACTGTATCGATTGGAA 1644
          || | ||  | ||  ||||||||||||||||||||||||||||||||||||||||||||||
Db    1438 AATAGCAATGAGACCAATATCACCATGAGTGCAGAGGTGGCAGAACTGTATCGATTGGAG 1497

Qy   1645 TTGGGAGATTATAAATTAGTAGAGATCACTCCAATTGGCTTGGCCCCCACAGATGTGAAG 1704
          ||||||||||  ||||||  |||||||  ||||||||||||||||||||||||  ||| |
Db    1498 TTGGGAGATTACAAATTAATAGAGATTACTCCAATTGGCTTGGCCCCCACAAGTGTAAGA 1557

Qy   1705 AGGTACACTACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCTTTGTGCTAGGGTTCTTG 1764
          ||||||||  ||    |||  ||||||||||||  ||||||||||||||||||||||||||||
Db    1558 AGGTACACCACAACTGGTGCCTCAAGAAATAAGAGAGGGGTCTTTGTGCTAGGGTTCTTG 1617

Qy   1765 GGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGTTGACGCTGACCGCTCAG 1824
          ||||||||||  ||  ||||||||||||||||||||||||  ||||||||  | |||||||
Db    1618 GGTTTTCTCGCGACAGCAGGTTCTGCAATGGGCGCGGCGTCCGTGACGCTGTCGGCTCAG 1677

Qy   1825 TCCCGAACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGCTGTTGGACGTGGTCAAG 1884
          |||||  |||||  ||||||||||||||||||||||||||||||||||||  |||||||||
Db    1678 TCCCGGACTTTGTTGGCTGGGATAGTGCAGCAACAGCAACAGCTGTTGGATGTGGTCAAG 1737

Qy   1885 AGACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACAAAGAACCTCCAGACTAGGGTC 1944
          ||||||||||||||||||||||||||||||||||||  ||||||||||||||||  |||
Db    1738 AGACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACTAAGAACCTCCAGACTAGAGTC 1797

Qy   1945 ACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATGCTTGGGGATGTGCGTTT 2004
          |||||  ||||||||||  | |||||  ||||||||||  |||  |  ||||||||||  |||
Db    1798 ACTGCTATCGAGAAGTACCTGAAGGATCAGGCGCAGCTAAATTCATGGGGATGTGCTTTT 1857

Qy   2005 AGACAAGTCTGCCACACTACTGTACCATGGCCAAATGCAAGTCTAACACCAAAGTGGAAC 2064
          ||  ||||||||  |||||||||||||||||||||||||||  ||    |  ||  ||  ||||||
```

```
Db   1858 AGGCAAGTCTGTCACACTACTGTACCATGGCCAAATGAAACATTGGTGCCTAATTGGAAC 1917

Qy   2065 AATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGGAAGAAAATATAACAGCC 2124
          |||  ||||||||||||||||||  || ||||||||||| | || | |||||||||
Db   1918 AATATGACTTGGCAAGAGTGGGAAAGACAGGTTGACTTCCTAGAGGCAAATATAACTCAA 1977

Qy   2125 CTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATGAATTACAAAAGTTGAAT 2184
            |  |||| || |||||||||||| ||||| ||||| |||||||||| ||||| || |||
Db   1978 TTATTAGAAGAAGCACAAATTCAGCAAGAAAAGAATATGTATGAATTGCAAAAATTAAAT 2037

Qy   2185 AGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGATAAAGTATATACAATAT 2244
          |||||||| | ||||||||||||||||||||| |||||||||||  ||||||||||||
Db   2038 AGCTGGGATATCTTTGGCAATTGGTTTGACCTTACTTCTTGGATAAGATATATACAATAT 2097

Qy   2245 GGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGATCTATATAGTACAAATG 2304
          || ||   ||||||| |||||||||||  |||||||||||| || ||| |||| ||||||
Db   2098 GGTGTACTTATAGTTCTAGGAGTAATAGGGTTAAGAATAGTAATATATGTAGTGCAAATG 2157

Qy   2305 CTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTCTCTTCCCCACCCTCTTATTTCCAG 2364
            |||||| |||||| ||||| |||||||||||||||||||||| ||| ||||| | |||
Db   2158 TTAGCTAGGTTAAGACAGGGTTATAGGCCAGTGTTCTCTTCCCCTCCCGCTTATGTTCAG 2217

Qy   2365 CAGACCCATATCCAACAGGACCCGGCACTGCCAACCAGAGAAGGCAAAGAAAGAGACGGT 2424
          |||| || |||||  ||| || || || |||||||| ||||||  ||||| ||||||||
Db   2218 CAGATCCCTATCCACAAGGGCCAGGAACCGCCAACCAAAGAAGGAGAAGAAGGAGACGGT 2277

Qy   2425 GGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAATATATTCATTTCCTGATC 2484
          |||||  | ||||||| ||| || ||||||||||||||||||||||||||||||||||||
Db   2278 GGAGACAGAGGTGGCAGCAGATCTTGGCCTTGGCAGATAGAATATATTCATTTCCTGATC 2337

Qy   2485 CGCCAACTGATACGCCTCTTGACTTGGCTATTCAGCAACTGCAGAACCTTGCTATCGAGA 2544
          |||||  |||||||||||||||||||||||||||||||| ||||||    | ||||| |||
Db   2338 CGCCAGTTGATACGCCTCTTGACTTGGCTATTCAGCAGCTGCAGGGATTGGCTATTGAGG 2397

Qy   2545 GTATACCAGATCCTCCAACCAATACTCCAGAGGCTCTCTGCGACCCTACAGAGGATTCGA 2604
             |||||||||||||||||| | |||||||| |||||  ||||  | || ||  | ||
Db   2398 AGCTACCAGATCCTCCAACCAGTGCTCCAGAGCCTCTCAACGACGTTGCAAAGAGTCCGT 2457

Qy   2605 GAAGTCCTCAGGACTGAACTGACCTACCTACAATATG 2641
          |||||| |||| | |||| |  ||||||||||||||
Db   2458 GAAGTCATCAGAATTGAAATAGCCTACCTACAATATG 2494


RESULT 9
AAN80859
ID   AAN80859 standard; cDNA; 9643 BP.
XX
AC   AAN80859;
XX
DT   15-NOV-1990  (first entry)
XX
DE   Sequence of entire HIV-2 ROD genome.
XX
KW   LAV-II ROD; ss.
XX
OS   Human immunodeficiency virus ROD.
XX
FH   Key             Location/Qualifiers
FT   CDS             545..2110
FT                   /*tag= a
FT                   /note="AAP80801"
FT   CDS             1828..4908
FT                   /*tag= b
FT                   /note="AAP81773"
FT   CDS             4841..5485
FT                   /*tag= c
FT                   /note="AAP81774"
FT   CDS             5316..5654
FT                   /*tag= d
FT                   /note="AAP81775"
FT   CDS             5654..5968
FT                   /*tag= e
FT                   /note="AAP81776"
FT   CDS             5817..6113
FT                   /*tag= f
FT                   /note="AAP81777"
FT   CDS             6043..6153
FT                   /*tag= g
FT                   /note="AAP81778"
FT   CDS             6119..8692
FT                   /*tag= h
FT                   /note="AAP81779"
FT   CDS             8220..8372
FT                   /*tag= i
FT                   /note="AAP81780"
FT   CDS             8242..8508
FT                   /*tag= j
FT                   /note="AAP81781"
FT   CDS             8529..9299
FT                   /*tag= k
FT                   /note="AAP81782"
XX
PN   WO8805440-A.
XX
PD   28-JUL-1988.
XX
PF   15-JAN-1988;   88WO-FR00025.
XX
PR   15-APR-1987;   87FR-0005398.
XX
PA   (INSP ) INST PASTEUR.
PA   (ALIZ/) ALIZON M.
XX
PI   Alizon M,  Montagnier L,  Guetard D,  Clavfl F,  Sonigo P,  Guyader M;
PI   Tiollais P,  Chakrabarti L,  Desrosiers R;
XX
DR   WPI; 1988-220290/31.
DR   P-PSDB; AAP80801, AAP81773, AAP81774, AAP81775, AAP81776, AAP81777,
DR   AAP81778, AAP81779, AAP81780, AAP81781, AAP81782.
XX
PT   New peptide(s) with immunological properties of HIV-2 envelope protein -
PT   having the structure of simian immune deficiency virus proteins,
PT   useful in diagnosis and of vaccine components
XX
PS   Disclosure; Fig 1A; 86pp; French.
XX
CC   The SQ in AAN80859 was deposited on 21/2/86 at the CNCM under number
CC   I-522, reference name LAV-II ROD. It is the cDNA to HIV-2 ROD genomic
CC   RNA. SQ was compared with the SQ of the genome of SIV (Mac) (AAN80860) to
CC   identify common regions.
XX
SQ   Sequence 9643 BP; 3296 A; 1969 C; 2399 G; 1979 T; 0 other;

  Query Match             53.5%;  Score 1412.4;  DB 9;  Length 9643;
  Best Local Similarity   73.9%;  Pred. No. 1.1e-314;
  Matches 1912;  Conservative    0;  Mismatches  616;  Indels   60;  Gaps    7;

Qy       2 CTAGAAGCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTGT 61
            |||||| |||| |||| |||||||||||||    || ||||||||||| ||| | |||
Db    5955 CTAGAAACATGCAATAACTCATGCTATTGTAAGCGATGCTGCTACCATTGTCAGATGTGT 6014

Qy      62 TTTCTTAAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCCG 121
           ||||| || || ||  | |||||||||||||| | |    | |   ||||||| |||||
Db    6015 TTTCTAAACAAGGGGCTCGGGATATGTTATGAACGAAAGGGCAGACGAAGAAGGACTCCA 6074

Qy     122 AAAAAGGCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGGA 181
           || ||  ||||| || || | ||| || || || |||||| ||||||          ||
Db    6075 AAGAAAACTAAGACTCATCCGTCTCCTACACCAGACAAGTGAGTATG---------ATGA 6125

Qy     182 ATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATATG 241
           |||||||||||||  ||||| ||  |    ||||  |       | ||||| || | |||||
Db    6126 ATCAGCTGCTTATTGCCATTTTATTAGCTAGTGCTTGCTTAGTATATTGCACCCAATATG 6185

Qy     242 TCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCAA 301
           | || || || ||||| |||||  | ||||  ||||| || ||||||||||||||||||
Db    6186 TAACTGTTTTCTATGGCGTACCCACGTGGAAAAATGCAACGATTCCCCTCTTTTGTGCAA 6245
```



```
          | |||| ||||| |||||| |||||||||||||||| ||||| ||||||||||  | |
Db    8364 CCTGGCCGATAGCATATATACATTTCCTGATCCGCCAGCTGATTCGCCTCTTGACCAGAC 8423

Qy    2513 TATTCAGCAACTGCAGAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCC 2572
           ||| ||||| ||||||   ||| ||||| ||    | || || |||||||  ||   ||
Db    8424 TATACAGCATCTGCAGGGACTTACTATCCAGGAGCTTCCTGACCCTCCAACTCATCTACC 8483

Qy    2573 AGAGGCTC 2580
           |||  |||
Db    8484 AGAATCTC 8491


RESULT 10
AAQ20616
ID   AAQ20616 standard; DNA; 9672 BP.
XX
AC   AAQ20616;
XX
DT   19-MAY-1992  (first entry)
XX
DE   ROD HIV-2 isolate complete genome.
XX
KW   Human immunodeficiency virus; AIDS; ss.
XX
OS   Human immunodeficiency virus-2 ROD isolate.
XX
FH   Key             Location/Qualifiers
FT   CDS             546..2111
FT                   /*tag= a
FT                   /note= "GAG gene"
FT   Misc_feature    546..950
FT                   /*tag= b
FT                   /note= "GAG p 16"
FT   Misc_feature    951..1640
FT                   /*tag= c
FT                   /note= "GAG p 26"
FT   Misc_feature    1701..2111
FT                   /*tag= d
FT                   /note= "GAG p 12"
FT   Misc_feature    1829..4936
FT                   /*tag= e
FT                   /note= "polymerase"
FT   Misc_feature    4869..5513
FT                   /*tag= f
FT                   /note= "Q protein"
FT   Misc_feature    5682..5996
FT                   /*tag= g
FT                   /note= "R protein"
FT   Misc_feature    5344..5679
FT                   /*tag= h
FT                   /note= "X protein"
FT   Misc_feature    5682..5996
FT                   /*tag= i
FT                   /note= "Y protein"
FT   Misc_feature    6147..8720
FT                   /*tag= j
FT                   /note= "Env protein"
FT   Misc_feature    8557..9324
FT                   /*tag= k
FT                   /note= "F protein"
FT   Exon            5845..6140
FT                   /*tag= l
FT                   /note= "TAT gene exon 1"
FT   Exon            8307..8400
FT                   /*tag= m
FT                   /note= "TAT gene exon 2"
FT   Exon            6071..6140
FT                   /*tag= n
FT                   /note= "ART gene exon 1"
FT   Exon            8307..8536
FT                   /*tag= o
FT                   /note= "ART gene exon 2"
FT   LTR             1..173
FT                   /*tag= p
FT                   /note= "LTR:R 1"
FT   LTR             9498..9671
FT                   /*tag= q
FT                   /note= "LTR:R 2"
FT   Misc_feature    174..299
FT                   /*tag= r
FT                   /note= "U5"
FT   Misc_feature    8942..9497
FT                   /*tag= s
FT                   /note= "U3"
XX
PN   US5079342-A.
XX
PD   07-JAN-1992.
XX
PF   11-FEB-1987;   87US-0U13477.
XX
PR   11-FEB-1987;   87US-0013477.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Alizon M,  Montagnier L,  Geutard D,  Clavel F,  Sonigo P,  Guyader M;
XX
DR   WPI; 1992-041067/05.
DR   P-PSDB; AAR20596, AAR20597, AAR20598, AAR20599, AAR20600, AAR20601,
DR   AAR20602, AAR20603, AAR20604, AAR20605, AAR20606, AAR20607.
XX
PT   Peptide(s) corresp. to HIV-2 amino acid sequences - used in
PT   diagnosis in vaccines and in prodn. of antibodies for diagnosis
XX
PS   Disclosure; Page 12; 30pp; English.
XX
CC   The sequence is that of the complete genomic sequence of the ROD HIV-2
CC   isolate. Its genes and their major expression products are indicated
CC   in the features. It was isolated from a CEM cell line infected with
CC   the ROD isolate.
XX
SQ   Sequence 9672 BP; 3314 A; 1981 C; 2394 G; 1983 T; 0 other;

  Query Match             53.2%;  Score 1404.8;  DB 13;  Length 9672;
  Best Local Similarity   73.7%;  Pred. No. 8.8e-313;
  Matches 1907;  Conservative    0;  Mismatches  622;  Indels   59;  Gaps    7;

Qy      2 CTAGAAGCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTGT 61
          |||||| ||||| ||||| |||||||||||||   || ||||||||||| ||| | |||
Db   5983 CTAGAAACATGCAATAACTCATGCTATTGTAAGCGATGCTGCTACCATTGTCAGATGTGT 6042

Qy     62 TTTCTTAAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCCG 121
          ||||| || || ||  | |||||||||||||| | |    | |   ||||||| |||||
Db   6043 TTTCTAAACAAGGGGCTCGGGATATGTTATGAACGAAAGGGCAGACGAAGAAGGACTCCA 6102

Qy    122 AAAAAGGCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGGA 181
          || ||  ||||| || || | ||| || || ||    |         |   || |   ||
Db   6103 AAGAAAACTAAGACTCATCCGTCTCCTACACCAGCACA--------AGTGAGTATGATGA 6154

Qy    182 ATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATATG 241
          |||||||||||||| ||||| ||  |    ||||  |       | ||||| || | |||||
Db   6155 ATCAGCTGCTTATTGCCATTTTATTAGCTAGTGCTTGCTTAGTATATTGCACCCAATATG 6214

Qy    242 TCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCAA 301
          | || || || ||||| |||||  | ||||  ||||| || ||||||||||||||||||||
Db   6215 TAACTGTTTTCTATGGCGTACCCACGTGGAAAAATGCAACCATTCCCCTCTTTTGTGCAA 6274

Qy    302 CCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCAG 361
          |||  |||||||||||||||||||| |  |||||| | || || |||| |||||||   |
Db   6275 CCAGAAATAGGGATACTTGGGGAACCATACAGTGCTTGCCTGACAATGATGATTATCAGG 6334

Qy    362 AAGTGGCCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAGG 421
          || |  |  | ||||| |||||    |||||||| |||||||||||||| |||||||| |
Db   6335 AAATAACTTTGAATGTAACAGAGGCTTTTGATGCATGGAATAATACAGTAACAGAACAAG 6394

Qy    422 CAATAGAGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCCC 481
```

```
Db   6395 CAATAGAAGATGTCTGGCATCTATTCGAGACATCAATAAAACCATGTGTCAAACTAACAC 6454

Qy    482 CATTATGCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAAT 541
Db   6455 CTTTATGTGTAGCAATGAAATGC------AGCAGCACAGAGAGCAGCACAGGGAACAACA 6508

Qy    542 CAATAACAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTCA 601
Db   6509 CAACCTCAAAGAGCACAAGCACAACCACAACCACACCCACAGACCAGGAGCAAGAGATAA 6568

Qy    602 ATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCAAATGA 661
Db   6569 GTGAGGATACTCCATGCGCACGCGCAGACAACTGCTCAGGATTGGGAGAGGAAGAAACGA 6628

Qy    662 TAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAATGAAA 721
Db   6629 TCAATTGCCAGTTCAATATGACAGGATTAGAAAGAGATAAGAAAAAACAGTATAATGAAA 6688

Qy    722 CTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGAT 781
Db   6689 CATGGTACTCAAAAGATGTGGTTTGTGAGACAAATAATAGCAC---AAATCAGACCCAGT 6745

Qy    782 GTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGGG 841
Db   6746 GTTACATGAACCATTGCAACACATCAGTCATCACAGAATCATGTGACAAGCACTATTGGG 6805

Qy    842 ATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGACA 901
Db   6806 ATGCTATAAGGTTTAGATACTGTGCACCACCGGGTTATGCCCTATTAAGATGTAATGATA 6865

Qy    902 CAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGGA 961
Db   6866 CCAATTATTCAGGCTTTGCACCCAACTGTTCTAAAGTAGTAGCTTCTACATGCACCAGGA 6925

Qy    962 TGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGAA 1021
Db   6926 TGATGGAAACGCAAACTTCCACATGGTTTGGCTTTAATGGCACTAGAGCAGAGAATAGAA 6985

Qy   1022 CTTATATTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTATA 1081
Db   6986 CATATATCTATTGGCATGGCAGAGATAATAGAACTATCATCAGCTTAAACAAATATTATA 7045

Qy   1082 ATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATGT 1141
Db   7046 ATCTCAGTTTGCATTGTAAGAGGCCAGGGAATAAGACAGTGAAACAAATAATGCTTATGT 7105

Qy   1142 CTGGATTGGTTTT------CCACTCACAACCAATCAATGATAGGCCAAAGCAGGCATGGT 1195
Db   7106 CAGGACATGTGTTTCACTCCCACTACCAGCCGATCAATAAAAGACCCAGACAAGCATGGT 7165

Qy   1196 GTTGGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAAC 1255
Db   7166 GCTGGTTCAAAGGCAAATGGAAAGACGCCATGCAGGAGGTGAAGGAAACCCTTGCAAAAC 7225

Qy   1256 ATCCCAGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGAGGAG 1315
Db   7226 ATCCCAGGTATAGAGGAACCAATGACACAAGGAATATTAGCTTTGCAGCGCCAGGAAAAG 7285

Qy   1316 G---AGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTA 1372
Db   7286 GCTCAGACCCAGAAGTAGCATACATGTGGACTAACTGCAGAGGAGAGTTTCTCTACTGCA 7345

Qy   1373 AAATGAATTGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGG 1432
Db   7346 ACATGACTTGGTTCCTCAATTGGATAGAGAATAAGACAC---------------------- 7384

Qy   1433 AACAGCATAAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATA 1492
Db   7385 ---------ACCGCAATTATGCACCGTGCCATATAAAGCAAATAATTAACACATGGCATA 7435

Qy   1493 AAGTAGGCAAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAG 1552
```

```
Db   7436 AGGTAGGGAGAAATGTATATTTGCCTCCCAGGGAAGGGGAGCTGTCCTGCAACTCAACAG 7495

Qy   1553 TGACCAGTCTCATAGCAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGA 1612
Db   7496 TAACCAGCATAATTGCTAACATTGACTGGCAAAACAATAATCAGACAAACATTACCTTTA 7555

Qy   1613 GTGCAGAGGTGGCAGAACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCA 1672
Db   7556 GTGCAGAGGTGGCAGAACTATACAGATTGGAGTTGGGAGATTATAAATTGGTAGAAATAA 7615

Qy   1673 CTCCAATTGGCTTGGCCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAA 1732
Db   7616 CACCAATTGGCTTCGCACCTACAAAAGAAAAAAGATACTCCTCTGCT---CACGGGAGAC 7672

Qy   1733 ATAAAAGAGGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAA 1792
Db   7673 ATACAAGAGGTGTGTTCGTGCTAGGGTTCTTGGGTTTTCTCGCAACAGCAGGTTCTGCAA 7732

Qy   1793 TGGGCGCGGCGTCGTTGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGC 1852
Db   7733 TGGGCGCGGCGTCCCTGACCGTGTCGGCTCAGTCCCGGACTTTACTGGCCGGGATAGTGC 7792

Qy   1853 AGCAACAGCAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCG 1912
Db   7793 AGCAACAGCAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAACTGTTGCGACTGACCG 7852

Qy   1913 TCTGGGGAACAAAGAACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACC 1972
Db   7853 TCTGGGGAACGAAAAACCTCCAGGCAAGAGTCACTGCTATAGAGAAGTACCTACAGGACC 7912

Qy   1973 AGGCGCAGCTGAATGCTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCAT 2032
Db   7913 AGGCGCGGCTAAATTCATGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCAT 7972

Qy   2033 GGCCAAATGCAAGTCTAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAA 2092
Db   7973 GGGTTAATGATTCCTTAGCACCTGACTGGGACAATATGACGTGGCAGGAATGGGAAAAAC 8032

Qy   2093 AGGTTGACTTCTTGGAAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAG 2152
Db   8033 AAGTCCGCTACCTGGAGGCAAATATCAGTAAAAGTTTAGAACAGGCACAAATTCAGCAAG 8092

Qy   2153 AGAAGAACATGTATGAATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTG 2212
Db   8093 AGAAAAATATGTATGAACTACAAAAATTAAATAGCTGGGATATTTTTGGCAATTGGTTTG 8152

Qy   2213 ACCTTGCTTCTTGGATAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATAC 2272
Db   8153 ACTTAACCTCCTGGGTCAAGTATATTCAATATGGAGTGCTTATAATAGTAGCAGTAATAG 8212

Qy   2273 TGTTAAGAATAGTGATCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGC 2332
Db   8213 CTTTAAGAATAGTGATATATGTAGTACAAATGTTAAGTAGGCTTAGAAAGGGCTATAGGC 8272

Qy   2333 CAGTGTTCTCTTCCCCACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCAC 2392
Db   8273 CTGTTTTCTCTTCCCCCCCCGGTTATATCCAACAGATCCATATCCACAAGGACCGGGGAC 8332

Qy   2393 TGCCAACCAGAGAAGGCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGC 2452
Db   8333 AGCCAGCCAACGAAGAAACAGAAGAAGACGGTGGAAGCAACGGTGGAGACAGATACTGGC 8392

Qy   2453 CTTGGCAGATAGAATATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGC 2512
Db   8393 CCTGGCCGATAGCATATATACATTTCCTGATCCGCCAGCTGATTCGCCTCTTGACCAGAC 8452

Qy   2513 TATTCAGCAACTGCAGAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCC 2572
Db   8453 TATACAGCATCTGCAGGGACTTACTATCCAGGAGCTTCCTGACCCTCCAACTCATCTACC 8512

Qy   2573 AGAGGCTC 2580
Db   8513 AGAATCTC 8520
```

```
RESULT 11
AAQ02830
ID   AAQ02830 standard; cDNA; 9360 BP.
XX
AC   AAQ02830;
XX
DT   29-MAY-1989  (first entry)
XX
DE   cDNA to HIV-2 RNA.
XX
KW   HIV; AIDS; Vaccine; pUC-HIV-2(GH-1).
XX
OS   HIV-2.
XX
FH   Key             Location/Qualifiers
FT   CDS             544..2109
FT                   /*tag= a
FT                   /label=Gag reading frame
FT   CDS             1830..4934
FT                   /*tag= b
FT                   /label=Pol gene reading frame.
FT   CDS             4867..5511
FT                   /*tag= c
FT                   /label=Q gene reading frame
FT   CDS             5342..5677
FT                   /*tag= d
FT                   /label=X gene reading frame
FT   CDS             5680..5994
FT                   /*tag= e
FT                   /label=R gene reading frame
FT   CDS             6148..8703
FT                   /*tag= f
FT                   /label=Env gene reading frame.
FT   CDS             8540..9304
FT                   /*tag= g
FT                   /label=F gene reading frame.
XX
PN   JP01289486-A.
XX
PD   21-NOV-1989.
XX
PF   16-MAY-1988;   88JP-0119024.
XX
PR   16-MAY-1988;   88JP-0119024.
XX
PA   (TOFU ) TOA NENRYO KOGYO KK.
PA   (FUJI-) FUJI KEBIO KK.
XX
DR   WPI; 1990-005177/01.
DR   P-PSDB; AAR04024-30.
XX
PT   DNA indicating complement to RNA gene -
PT   of Human Immunodeficiency Virus type 2 used for new vaccine or
PT   diagnostic for AIDS virus.
XX
PS   Claim 2; Fig.4; 12pp; Japanese.
XX
CC   cDNA to novel HIV-2 (GH-1) has been integrated into plasmid
CC   pUC HIV-2(GH-1). Useful for diagnosis and vaccination against the virrus.
CC   Carries 7 overlapping genes in varrious reading frames, includding Gag,
CC   Pol and Env.
XX
SQ   Sequence 9360 BP; 3211 A; 1944 C; 2317 G; 1888 T; 0 other;

  Query Match           52.8%;  Score 1395.6;  DB 11;  Length 9360;
  Best Local Similarity  73.1%;  Pred. No. 1.1e-310;
  Matches 1905;  Conservative    0;  Mismatches  629;  Indels   72;  Gaps    6;

Qy       2 CTAGAAGCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTGT 61
Db    5981 CTAGAGGCATGTACTAACTCATGCTATTGTAAGCAGTGCAGCTTCCATTGCCAGCTGTGT 6040

Qy      62 TTTCTTAAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCCG 121
Db    6041 TTCTTGAAAAAAGGGCTCGGGATATGGTATGCACGAAAAAGCAGACGGAGAAGAACTCCA 6100

Qy     122 AAAAAGGCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGGA 181
Db    6101 AGAAAGACTAAGACTCATTCGTCTTCTGCATCAGACAAGTGAGTAAGATGTGTGGTAAGA 6160

Qy     182 ATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATATG 241
Db    6161 GTCTACTATGTGTTGCCAGCTTGCTAGCTAGTGCTTACTTAGTATATTGCACCCAATATG 6220

Qy     242 TCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCAA 301
Db    6221 TGACTGTTTTCTATGGCGTGCCCGTGTGGAGAAATGCATCCATTCCCCTCTTTTGTGCAA 6280

Qy     302 CCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCAG 361
Db    6281 CTAAAAACAGAGATACTTGGGGAACCATACAGTGCAAGCCAGACAATGATGATTATCAAG 6340

Qy     362 AAGTGGCCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAGG 421
Db    6341 AAATAACCTTGAATGTGACAGAGGCTTTCGACGCATGGGATAATACAGTAACAGAACAAG 6400

Qy     422 CAATAGAGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCCC 481
Db    6401 CAGTAGAAGATGTCTGGAGTCTATTTGAGACATCAATAAAACCATGCGTCAAACTAACAC 6460

Qy     482 CATTATGCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAAT 541
Db    6461 CTCTATGTGTAGCAATGAGTTGTAACA--------------------GCACCACTAACA 6499

Qy     542 CAATAACAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTCA 601
Db    6500 ACACAACCACAACAGGGAGCACAACAGGGATGTCAGAGATAAATGAA------------- 6546

Qy     602 ATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCAAATGA 661
Db    6547 --------ACTTCTCCATCATACTCAGACAACTGCACAGGATTGGGGAAGGAAGAGATAG 6598

Qy     662 TAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAATGAAA 721
Db    6599 TCAATTGTCAGTTCTATATGACAGGATTAGAGAGAGATAAGAAAAAACAGTATAATGAAA 6658

Qy     722 CTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGAT 781
Db    6659 CATGGTACTCAAAAGATGTAGTTTGTGAATCAAATAATACCAAAGATGGGAAAAACAGAT 6718

Qy     782 GTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGGG 841
Db    6719 GTTACATGAACCATTGCAACACATCAGTCATCACAGAGTCATGTGATAAGCACTATTGGG 6778

Qy     842 ATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGACA 901
Db    6779 ATGCTATAAAGTTTAGATACTGTGCACCACCGGGTTATGCCCTACTAAGATGCAATGATA 6838

Qy     902 CAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGGA 961
Db    6839 CCAATTATTCAGGCTTCGAGCCAAAGTGTTCTAAAGTAGTAGCTTCTACATGCACAAGGA 6898

Qy     962 TGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGAA 1021
Db    6899 TGATGGAAACGCAAACCTCCACTTGGTTTGGCTTTAATGGCACTAGGGCAGAGAATAGAA 6958

Qy    1022 CTTATATTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTATA 1081
Db    6959 CATATATCTATTGGCATGGCAGAGATAATAGGACTATCATCAGCTTAAACAAGTATTATA 7018

Qy    1082 ATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATGT 1141
Db    7019 ACCTTTCTATACATTGTAAGAGACCAGGAAACAAGACAGTTGTACCAATAACACTCATGT 7078
```

```
Qy   1142 CTGGATTGGTTTTCCACTCACAACCAATCAATGATAGGCCAAAGCAGGCATGGTGTTGGT 1201
          | || || || || ||||| || |||||||||    || || | ||| ||||||||| ||||
Db   7079 CAGGGTTAGTGTTTCACTCCCAGCCAATCAATACAAGACCTAGGCAAGCATGGTGCTGGT 7138

Qy   1202 TTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCCA 1261
          |   ||| || |||| ||| || ||  | ||||| |||| ||| || | ||||||||||
Db   7139 TCAAAGGCAAGTGGAGGGAAGCCATGCAGGAGGTAAAGCAAACCCTTATAAAACATCCCA 7198

Qy   1262 GGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGAGGAGG---AG 1318
          |||||| ||||| ||| | ||  | || || || || ||    || ||| ||||   ||
Db   7199 GGTATAAAGGAACCAACGACACAAAGAATATTAACTTTACAAAACCAGGAAGAGGCTCAG 7258

Qy   1319 ATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATGA 1378
          | || |||||  | | |||||||| || |||||||||| || |||||||| || ||||
Db   7259 ACCCAGAAGTGGCATACATGTGGACTAACTGCAGAGGAGAATTTCTCTACTGCAACATGA 7318

Qy   1379 ATTGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGGAACAGC 1438
            |||||| || |||||||||||| | ||                | | | |||
Db   7319 CTTGGTTCCTCAATTGGGTAGAAAACAG--------------------ACCGAATCAGA 7357

Qy   1439 ATAAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAG 1498
             |    ||||| | ||| || ||||| || |||||||| || || ||||| || ||||
Db   7358 CACAGCACAATTATGCGCCGTGCCATATAAGGCAAATAATTAATACCTGGCACAAGGTAG 7417

Qy   1499 GCAAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACCA 1558
          | |||||||| |||||||||| ||||| ||| |  | || || ||||| ||||||||||
Db   7418 GGAAAAATGTATATTTGCCTCCTAGAGAAGGACAGTTGACCTGCAACTCAACAGTGACCA 7477

Qy   1559 GTCTCATAGCAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTGCAG 1618
          |  | || || ||||| ||     |  || | ||||| || ||||| ||| | |||||||
Db   7478 GCATAATTGCTAACATTGA---CGTAAATAGTAACCAGACAAATATTACCTTTAGTGCAG 7534

Qy   1619 AGGTGGCAGAACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTCCAA 1678
          |||||||||||||| || ||||| ||||||||||||||||||||| ||||  | || ||||
Db   7535 AGGTGGCAGAACTATACCGATTAGAATTGGGAGATTATAAATTAATAGAAGTAACACCAA 7594

Qy   1679 TTGGCTTGGCCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAA 1738
          ||||||| || || ||    | ||| || || |  | | |    |    || || ||||
Db   7595 TTGGCTTCGCACCTACGAGAGAGAAAAGATATTCCTCGGCTC---CAGTGAGGAACAAAA 7651

Qy   1739 GAGGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCG 1798
          |||| || ||||||||||||||||||||||||||||| || ||||||||||||||||||
Db   7652 GAGGTGTGTTTGTGCTAGGGTTCTTGGGTTTTCTCGCGACAGCAGGTTCTGCAATGGGCG 7711

Qy   1799 CGGCGTCGTTGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGCAAC 1858
          |||||||  |||||||| | |||||||||| |||||| |||| ||||||||||||||||
Db   7712 CGGCGTCCCTGACGCTGTCGGCTCAGTCCCGGACTTTACTGGCCGGGATAGTGCAGCAAC 7771

Qy   1859 AGCAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCTGGG 1918
          |||||||||||||||||| |||||||||||||||||| ||||||||||||||||||||
Db   7772 AGCAACAGCTGTTGGACGTAGTCAAGAGACAACAAGAAATGTTGCGACTGACCGTCTGGG 7831

Qy   1919 GAACAAAGAACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGGCGC 1978
          |||| || || |||||| | || |||||||| |||||||||||||||||||||||||| |
Db   7832 GAACGAAAAATCTCCAGGCAAGAGTCACTGCTATCGAGAAGTACTTAAAGGACCAGGCAC 7891

Qy   1979 AGCTGAATGCTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGCCAA 2038
          || | ||| | ||||| |||||||||| || ||||||||||||||||||||||||  ||
Db   7892 AGTTAAATTCATGGGGGTGTGCGTTTAGGCAGGTCTGCCACACTACTGTACCATGGGTAA 7951

Qy   2039 ATGCAAGTCTAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTG 2098
          |||    | || | ||  | |||||||||  ||| ||||| || |||||   | | ||
Db   7952 ATGATTCTTTATCGCCTGACTGGAACAATATGACATGGCAGGAATGGGAAAAACAAGTCC 8011

Qy   2099 ACTTCTTGGAAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGA 2158
            || | | || | |||||| |          ||||  |||||||||||||||||| ||||
Db   8012 GCTACCTAGAGGCAAATATCAGTCAAAGTTTAGAACAGGCACAAATTCAACAAGAAAAGA 8071

Qy   2159 ACATGTATGAATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTG 2218
          | |||||||||||||||| | |||||||||||||| ||||||| ||||||||   |
Db   8072 ATATGTATGAATTACAAAAGCTAAATAGCTGGGATGTCTTTGGCAACTGGTTTGATTTAA 8131

Qy   2219 CTTCTTGGATAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAA 2278
          | || |||||| |||||||| |||||||||||||||||||| |||||||| ||| | ||||
Db   8132 CCTCCTGGATCAAGTATATTCAATATGGAGTTTATATAGTAGTAGGAGTGATAGTTTTAA 8191

Qy   2279 GAATAGTGATCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGT 2338
          ||||||  || |||||||| ||| || ||  | ||  |||| |||||||| || |
Db   8192 GAATAGCAATATATATAGTGCAATTGTTAAGTAGACTTAGAAAGGGCTATAGGCCTGTTT 8251

Qy   2339 TCTCTTCCCCACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAA 2398
          |||||||||| |||  |||| |||| |||| |||||||||   ||||| || || ||||
Db   8252 TCTCTTCCCCTCCCGGTTATCTCCAACAGATCCATATCCACACGGACCGGGGACAGCCAG 8311

Qy   2399 CCAGAGAAGGCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGC 2458
          |||  ||||| | ||||  |||||   |||| | || |||  ||  ||  | |||||||||
Db   8312 CCAACGAAGGAACAGAAGAAGACGACAGAGACGACGATGGTTACGACTTGTAGCCTTGGC 8371

Qy   2459 AGATAGAATATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCA 2518
            |||| | ||||| |||||||||||| || |||||  |||||||||||  |||||| ||
Db   8372 CGATAAACTATATACATTTCCTGATCCACCTACTGACTCGCCTCTTGACCGGGCTATACA 8431

Qy   2519 GCAACTGCAGAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGC 2578
            | |||||||  ||| ||||| |     | || || ||.|| ||  ||  |||||  |
Db   8432 AGATCTGCAGAGACTTACTATCCACGAACTCCCCGACCCACCGACTGATCTCCCAGAATC 8491

Qy   2579 TCTCTGCGACCCTACAGAGGATTCGA 2604
          |  | || | |    |  || |  ||
Db   8492 TAACAGCAATCAGGGACTGGCTGAGA 8517
```

RESULT 12
AAN80890
ID AAN80890 standard; cDNA; 9633 BP.
XX
AC AAN80890;
XX
DT 07-NOV-1990 (first entry)
XX
DE Sequence of cDNA clone HIV-2 SBL/ISY of HIV related retrovirus strain.
XX
KW HIV vaccine; HIV strain SBL-6669-85; ss.
XX
OS Human immunodeficiency virus.
XX
FH Key Location/Qualifiers
FT CDS 1..9633
FT /*tag= a
FT /note="SQ in AAP81769"
FT CDS 2..9631
FT /*tag= b
FT /note="SQ in AAP81770"
FT CDS 3..9632
FT /*tag= c
FT /note="SQ in AAP81771"
XX
PN WO8808449-A.
XX
PD 03-NOV-1988.
XX
PF 28-APR-1988; 88WO-SE00218.
XX
PR 28-APR-1987; 87SE-0001765.
XX
PA (SBLS-) SBL STATENS BEKTERI.
XX
PI Albert J, Biberfeld G, Fenyo EM, Norrby E;
XX
DR WPI; 1988-322769/45.
DR P-PSDB; AAP81769, AAP81770, AAP81771.
XX
PT HIV related human retro-virus strain -
PT used for obtaining antigens for assays and vaccines and for
PT prodn. of antibodies for assays
XX
PS Claim 7; Fig 4; 28pp; English.

XX
CC cDNA clone (HIV-2 SBL/ISY) characterised by having the primary nucleotide
CC sequence in n80890 or a degenerate or a part thereof, esp. a clone having
CC capacity to give prodn. of infectious virus particles is claimed.
CC HIV-2 SBL/ISY represents the complete genome of the virus SBL-6699
CC (=SBL-6689-85). The proviral DNA was obtd. from a genomic library
CC constructed from HUT-78 cells infected with SBL-6669-85 virus using
CC the lambda-phage vector EMBL-3. SBL-6669-85 was isolated from lymphocytes
CC of a West African woman. Protection is requested for the entire genome
CC disclosed in n80890 and for parts thereof, and corresp. to various genes
CC such as the gag gene (corresp. to nucleotides 547 to 2106), the pol gene
CC (nucleotides 1827-4931) and the env gene (nucleotides 6144 to 8682), the
CC corresp. AA sequences and parts thereof and various products derived
CC therefrom, or use thereof, such as clones prepd. by recombinant vector
CC method, HIV test devices and methods.
XX
SQ Sequence 9633 BP; 3288 A; 1943 C; 2394 G; 2008 T; 0 other;

Query Match 52.1%; Score 1377.2; DB 9; Length 9633;
Best Local Similarity 73.0%; Pred. No. 1.9e-306;
Matches 1886; Conservative 0; Mismatches 623; Indels 75; Gaps 6;

```
Qy      2 CTAGAAGCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTGT 61
          |||||| ||||| |||||| ||||| ||||||   ||| |||| |||||||||| | |||
Db   5975 CTAGAAACATGCAATAACAAATGCTTTTGTAAGGGGTGCTGCTTCCATTGCCAGCTGTGT 6034

Qy     62 TTTCTTAAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCCG 121
          ||| | || || ||  | ||||||||||||| | |    | |   ||||||| | ||||
Db   6035 TTTTTAAACAAGGGGCTCGGGATATGTTATGACCGAAAGGGCAGACGAAGAAGGAGTCCG 6094

Qy    122 AAAAAGGCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGGA 181
          || ||  |||||||| || | ||| |||||||| |||||| ||||      ||  |   |
Db   6095 AAGAAAACTAAGGCTCATTCGTCTCCTGCATCAGACAAGTGAGTACAATGAGTGGTAAAA 6154

Qy    182 ATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATATG 241
           |||||||||| | ||| |  ||||    ||||  |     || ||||| ||   |||||
Db   6155 TTCAGCTGCTTGTTGCCTTTCTGCTAACTAGTGCTTGCTTAATATATTGCACCAAATATG 6214

Qy    242 TCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCAA 301
          | || || || ||||| ||||| |  ||||  |||||  | ||||||||||||||||||
Db   6215 TGACTGTTTTCTATGGAGTACCCGTGTGGAAAAATGCATCCATTCCCCTCTTTTGTGCAA 6274

Qy    302 CCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCAG 361
          | || ||||| ||||||||||||||| |  |||||| | |||||  |||| |||||||  ||
Db   6275 CTAAAAATAGAGATACTTGGGGAACCATACAGTGCTTGCCAGACAATGATGATTATCAAG 6334

Qy    362 AAGTGGCCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAGG 421
          |  |  |  | ||||| |||||    ||||| || ||| |||||| ||| ||||||||  |
Db   6335 AGATACCTTTGAATGTAACAGAGGCTTTTGACGCATGGGATAATATAGTAACAGAACAAG 6394

Qy    422 CAATAGAGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCCC 481
          || |||| ||||| ||| | || |||||||| |||||||| || ||||| ||||||  | |
Db   6395 CAGTAGAAGATGTCTGGAATCTATTTGAGACATCAATAAAACCATGTGTCAAATTAACGC 6454

Qy    482 CATTATGCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAAT 541
          | |||||  | || ||||  || ||   ||||
Db   6455 CTTTATGTGTAACAATGAACTGTAACGCAAGT---------------------------- 6486

Qy    542 CAATAACAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTCA 601
                |||||      ||   |||| |||    || |   |   || ||| | |
Db   6487 -----------ACAGAGAGCGCAGTTGCAACTACAAGCCCATCTGGACCTGATATGATAA 6535

Qy    602 ATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCAAATGA 661
          |||| ||| | | || |||     | | |||||| |||  ||    | ||  | |||
Db   6536 ATGATACTGATCCATGCATACAATTGAACAATTGCTCAGGACTGAGGGAGGAAGACATGG 6595

Qy    662 TAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAATGAAA 721
          |    ||| | ||||| ||||||||  || |   ||| |||||||| |||| | |||||
Db   6596 TCGAGTGTCAGTTCAATATGACAGGACTAGAGTTAGATAAGAAAAAACAGTATAGTGAAA 6655

Qy    722 CTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGAT 781
          | ||||||||   |||| |||| ||||||  ||  || | ||| |   |   ||  ||||
Db   6656 CCTGGTACTCAAAAGATGTGGTTTGTGAATCAGATAACAGCACAG---ACCGAAAAAGAT 6712

Qy    782 GTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGGG 841
          |||||||||||||| || ||||| || || |||  |||||| |||||||| || |||||||
Db   6713 GTTACATGAACCATTGCAACACATCAGTCATCACAGAGTCATGTGACAAGCACTATTGGG 6772

Qy    842 ATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGACA 901
          ||||||| |||||||| || |||||||| || |||| ||  ||||| || || |||||  |
Db   6773 ATGCTATGAGATTTAGATACTGTGCACCACCGGGTTTTGTCTTGCTAAGGTGCAATGATA 6832

Qy    902 CAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGGA 961
          | ||||| |||||||||  ||| || || ||||| || || |  |||| |||| ||||| |
Db   6833 CCAATTACTCAGGCTTTGAGCCCAATTGCTCTAAAGTAGTAGCTTCTACATGTACAAGAA 6892

Qy    962 TGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGAA 1021
          |||||||  || ||  |||||||||||  |||||||||||||  ||||| ||||||||||||
Db   6893 TGATGGAAACGCAACCTTCTACTTGGCTTGGCTTTAATGGCACTAGGGCAGAAAATAGAA 6952

Qy   1022 CTTATATTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTATA 1081
          | ||||| || |||||||||||||||||| || ||||| || || ||||| || |||||||
Db   6953 CATATATCTATTGGCATGGTAGGGATAACAGAACTATTATCAGCTTAAACAAATATTATA 7012

Qy   1082 ATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATGT 1141
          |||| || ||    ||||| ||||||| |||||| |||||| ||||| | ||  | ||||
Db   7013 ATCTCACCATACTTTGTAGGAGACCAGAAAATAAAACAGTTGTACCAATAACACTCATGT 7072

Qy   1142 CTGGATTGGTTTTCCACTCAC---AACCAATCAATGATAGGCCAAAGCAGGCATGGTGTT 1198
          | ||       || ||||| |   |    |||||| | |  || | ||| ||||||||
Db   7073 CAGGCCGCAGATTTCACTCCCAGAAGATCATCAATAAAAAACCCAGGCAAGCATGGTGCC 7132

Qy   1199 GGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATC 1258
          ||||   ||| | |||| ||| || ||  | |||||||| || ||| |||| |||||||
Db   7133 GGTTCAAAGGCGAGTGGAGGGAAGCCATGCAGGAGGTGAAACAAACCCTTGTAAAACATC 7192

Qy   1259 CCAGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGAGGAG--- 1315
          |||||||||  |||| ||  | ||  ||||||| || || || || || | |  ||
Db   7193 CCAGGTATAAAGGAACCAATGACACAAATAAAATTAACTTTACAGCACCAGAAAAAGACT 7252

Qy   1316 GAGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAA 1375
          ||| || ||||  | |  |||||||| || |||||||||| |||||||| || || |
Db   7253 CAGACCCAGAAGTAGCATATATGTGGACTAACTGCAGAGGAGAATTCCTCTATTGCAACA 7312

Qy   1376 TGAATTGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGGAAC 1435
          ||| |||||| || |||||||||||| | | ||          | || ||
Db   7313 TGACTTGGTTCCTTAATTGGGTAGAAAACAAGACGGGTCAACAGCATAAC---------- 7362

Qy   1436 AGCATAAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAG 1495
                    || ||||| || |||||    |||||||| || || |||||||| |
Db   7363 --------------TATGTGCCGTGCCATATAGAGCAAATAATTAATACCTGGCATAAGG 7408

Qy   1496 TAGGCAAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGA 1555
          |||| |||||||| |||||||||| || || ||||| |  | ||  | || |||||||
Db   7409 TAGGGAAAAATGTATATTTGCCTCCTAGGGAAGGAGAGTTGTCCTGCGAATCAACAGTGA 7468

Qy   1556 CCAGTCTCATAGCAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTG 1615
          ||||| |||| || ||||| |||      || |  |||  || ||||| ||| | ||||
Db   7469 CCAGTATCATTGCTAACATTGATGTTGATGGAGATAACCGGACAAATATTACCTTTAGTG 7528

Qy   1616 CAGAGGTGGCAGAACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTC 1675
          |||||||||||||||| || |||||||||||||| ||||||||||||||||  | || |
Db   7529 CAGAGGTGGCAGAACTATACCGATTGGAATTGGGGGATTATAAATTAGTAGAAGTAACAC 7588

Qy   1676 CAATTGGCTTGGCCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATA 1735
          |||||||||| |||| |||| |  || || ||| |  ||| |    |   ||| |||
Db   7589 CAATTGGCTTCGCCCCTACAGCAGAAAAAAGATACTCCTCTGCT---CCAGGGAGACATA 7645

Qy   1736 AAAGAGGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGG 1795
          | ||||| ||  |||||||||||||| | |||||||| | || |||||| ||||||||||
Db   7646 AGAGAGGTGTGCTTGTGCTAGGGTTCCTAGGTTTTCTCACGACAGCAGGTGCTGCAATGG 7705

Qy   1796 GCGCGGCGTCGTTGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGC 1855
          |||||  |||  |||||||| | |||||||| || ||||||||   ||||||||||||||
Db   7706 GCGCGCGGTCTCTGACGCTGTCGGCTCAGTCTCGGACTTTATTCCGTGGGATAGTGCAGC 7765
```

```
Qy   1856 AACAGCAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCT 1915
          ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   7766 AACAGCAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAAATGTTGCGACTGACCGTCT 7825

Qy   1916 GGGGAACAAAGAACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGG 1975
          ||||||| || ||||||||| | || |||||||| || ||||||||| ||   |||||||
Db   7826 GGGGAACTAAAAACCTCCAGGCAAGAGTCACTGCTATTGAGAAGTACCTAGCAGACCAGG 7885

Qy   1976 CGCAGCTGAATGCTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGC 2035
          |||  || ||| | ||||||||||||||||||||||||||||||||||||||||||||||
Db   7886 CGCGACTAAATTCATGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGG 7945

Qy   2036 CAAATGCAAGTCTAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGG 2095
          |||||  |   |||||||  ||||||||||   ||| ||||||| ||||| |  ||
Db   7946 TAAATGACACCTTAACACCTGAGTGGAACAACATGACATGGCAAGAATGGGAACACAAAA 8005

Qy   2096 TTGACTTCTTGGAAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGA 2155
          |  |||| | || | |||||| |  |      |||| ||||||||| || ||||| |
Db   8006 TCCGCTTCCTAGAGGCAAATATCAGTGAGAGTTTAGAACAGGCACAAATCCAGCAAGAAA 8065

Qy   2156 AGAACATGTATGAATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACC 2215
          |||| ||||||||  | ||||||| | |||||||||||||| ||||||||||||||||||
Db   8066 AGAATATGTATGAGCTGCAAAAGCTAAATAGCTGGGATGTTTTTGGCAATTGGTTTGACT 8125

Qy   2216 TTGCTTCTTGGATAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGT 2275
          |  | || ||||| |||||||| ||||||||||    ||||| |||||| || ||
Db   8126 TAACCTCCTGGATCAAGTATATTCAATATGGAGTCATGATAGTAGTAGGAATAGTAGCTC 8185

Qy   2276 TAAGAATAGTGATCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAG 2335
          | |||||||| || ||| |||||||||||||  ||   | ||  |||| |||||||| |
Db   8186 TCAGAATAGTAATATATGTAGTACAAATGCTAAGTAGACTTAGAAAGGGCTATAGGCCTG 8245

Qy   2336 TGTTCTCTTCCCCACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGC 2395
          | |||||||||| |||  |||| | || |||| |||||||||  |||||  || || ||
Db   8246 TTTTCTCTTCCCCCCCCGGTTATATTCAACAGATCCATATCCACAAGGACTGGGAACAGC 8305

Qy   2396 CAACCAGAGAAGGCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTT 2455
          ||  |||||||  | ||||  ||||| ||||  | | || ||| | ||| ||||||||||
Db   8306 CAGACAGAGAAGAAACAGAAGAAGACGTTGGGAACGACGTTGGAAGCAGATCCTGGCCTT 8365

Qy   2456 GGCAGATAGAATATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTAT 2515
          ||| |||||||||||| ||||||||||||||||  ||||| ||||||||||  | ||||
Db   8366 GGCCGATAGAATATATACATTTCCTGATCCGCCTGCTGATCCGCCTCTTGACCAGACTAT 8425

Qy   2516 TCAGCAACTGCAGAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGA 2575
           || || ||||||   ||| ||||| ||| | |||| ||||||||||||| |     |
Db   8426 ACAACAGCTGCAGGGACTTACTATCCAGACTCTACCTGATCCTCCAACCACTCAGAGACT 8485

Qy   2576 GGCT 2579
          ||||
Db   8486 GGCT 8489


RESULT 13
AAN92768
ID   AAN92768 standard; DNA; 9473 BP.
XX
AC   AAN92768;
XX
DT   18-MAY-1990  (first entry)
XX
DE   HIV-2 variant HIV-D194 clone.
XX
KW   HIV-2; AIDS; HIV-D194; ss;
XX
OS   HIV-2.
XX
PN   EP347365-A.
XX
PD   20-DEC-1989.
XX
PF   14-JUN-1988;   88EP-0710057.
XX
PR   14-JUN-1988;   88DE-3820223.
XX
PA   (CHEM ) CHEMOTHERAPEUT GEORG-SPE.
PA   (DIAG ) DIAGEN.
XX
PI   Henco K,  von Briesen H,  Immelmann A,  Kuhnel H,  Dietrich U;
PI   Rumsamen -,  Waigmann H,  Adamski M;
XX
DR   WPI; 1989-372573/51.
XX
PT   New HIV-2 virus variant isolates - used to obtain prods. for identifying
PT   infections by HIV-2 or vaccines for protection against AIDS infections.
XX
PS   Disclosure; fig.4; 27pp; English.
XX
CC   The sequence characterises the clone HIV-D194, it starts with R/U5 region
CC   of the LTR and ends with the U5 region. The DNA can be used to raise Abs
CC   for serological detection, and polypeptides encoded by the sequence used
CC   in vaccination.
XX
SQ   Sequence 9473 BP; 3215 A; 1953 C; 2374 G; 1923 T; 8 other;

  Query Match           51.4%;  Score 1356.4;  DB 10;  Length 9473;
  Best Local Similarity  72.4%;  Pred. No. 1.2e-301;
  Matches 1889;  Conservative    0;  Mismatches  636;  Indels   84;  Gaps    7;

Qy      2 CTAGAAGCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTGT 61
          ||||| ||||||  |||| ||||||||||||  ||||  | ||||||||||| | |||
Db   5982 CTAGAGGCATGCACTAACTCATGCTATTGTAAGCAGTGCAGTTACCATTGCCAGCTGTGT 6041

Qy     62 TTTCTTAAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCCG 121
          ||  | ||||||||  | ||||||| |||| || |    | |  ||||||| |||||
Db   6042 TTCTTGAAAAAAGGGCTCGGGATATGGTATGCGCGACAGGGCAGACGAAGAAGGACTCCA 6101

Qy    122 AAAAAGGCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGGA 181
          | |||  ||||| || || |  || ||||||||'| ||||||||||||    |  | |||
Db   6102 AGAAAAACTAAGACTCATCCGCCTCCTGCATCAGATAAGTAAGTATGGAGCCTGGTAGGA 6161

Qy    182 ATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATATG 241
          |||||||||||| | ||||| ||  |    ||||  |     || |||||  |  | |||||
Db   6162 ATCAGCTGCTTGTTGCCATTTTATTAACTAGTGCTTGCTTAATATATTGCAAACAATATG 6221

Qy    242 TCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCAA 301
          | || || || |||||  |||| || ||||| |||||  | ||||||||| ||||||||||
Db   6222 TGACTGTTTTCTATGGCATACCCGCGTGGAGAAATGCATCTATTCCCCTATTTTGTGCAA 6281

Qy    302 CCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCAG 361
          |||| ||||| ||||||||||| || |  |||||| | ||||| |||| ||||||   |
Db   6282 CCAAAAATAGAGATACTTGGGGGACCATCCAGTGCTTGCCAGACAATGATGATTATCAGG 6341

Qy    362 AAGTGGCCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAGG 421
          || |  || | ||||| ||||||   |||||||| ||| |||||||||| |||||||| |
Db   6342 AAATAACCTTAAATGTGACAGAAGCTTTTGATGCATGGGATAATACAGTAACAGAACAAG 6401

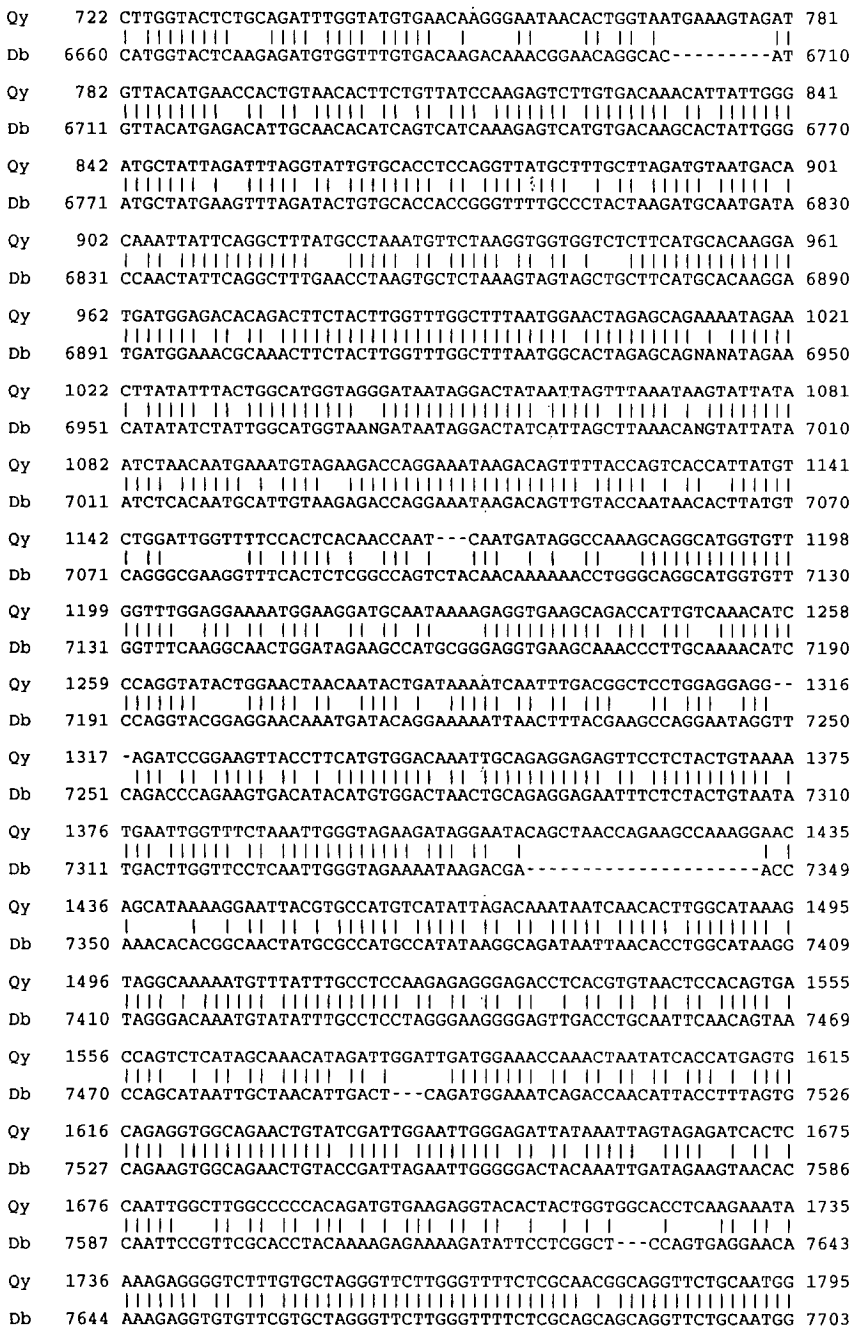
Qy    422 CAATAGAGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCCC 481
          ||||||| ||||| ||| ||| |||||||| |||||||| || ||||| || || | |
Db   6402 CAATAGAAGATGTCTGGAGACTGTTTGAGACATCAATAAAACCATGTGTCAAGTTGACGC 6461

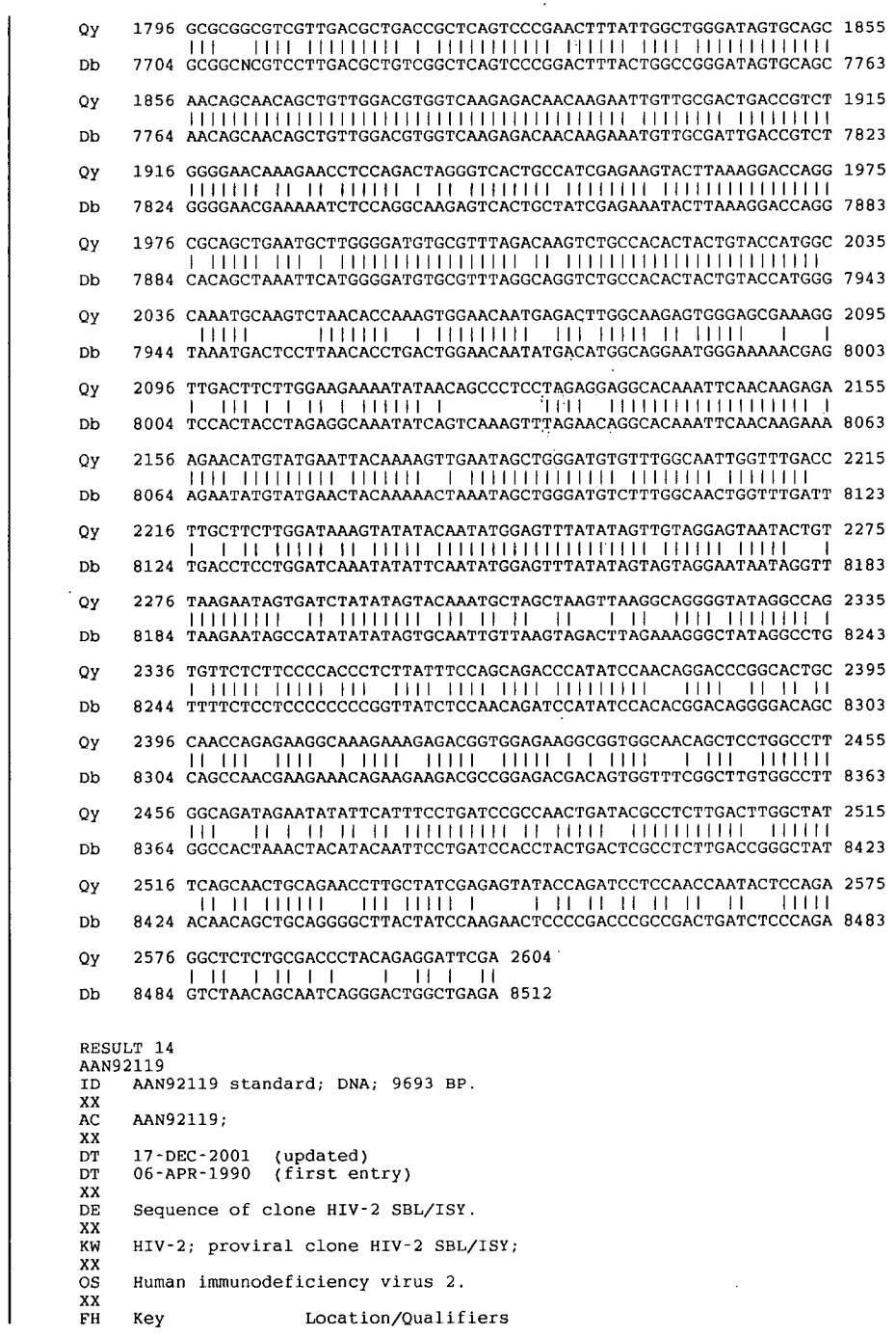
Qy    482 CATTATGCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAAT 541
          |  |||| |  | ||||  || |||
Db   6462 CCCTATGTGTGGCGATGAATTGTAAT---------------------------------- 6487

Qy    542 CAATAACAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTCA 601
                | |||  ||   || ||  | ||  |     || ||||  |    ||  |
Db   6488 --------ATAACTTCAGGGACTACCGCGACCCCGAGTCCACCAAACATTACAATAATAG 6539

Qy    602 ATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCAAATGA 661
          |||| | |  | | ||||||| | |  | || |||||||| || |   |||  || ||
Db   6540 ATGAAAATTCTACCTGTATAGGCGACAACAACTGCACAGGATTAGGGAAAGAAGAGGTGG 6599

Qy    662 TAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAATGAAA 721
          |    ||| | ||||| ||||| ||| || ||  ||| |||||||   |||| |||||
Db   6600 TTGAGTGTGAGTTCAATATGACGGGGCTAGAACAAGATAAGAAAAGGAAGTATAATGACG 6659
```





```
RESULT 14
AAN92119
ID   AAN92119 standard; DNA; 9693 BP.
XX
AC   AAN92119;
XX
DT   17-DEC-2001  (updated)
DT   06-APR-1990  (first entry)
XX
DE   Sequence of clone HIV-2 SBL/ISY.
XX
KW   HIV-2; proviral clone HIV-2 SBL/ISY;
XX
OS   Human immunodeficiency virus 2.
XX
FH   Key             Location/Qualifiers
```

```
FT   CDS             1..9693
FT                   /*tag=  a
FT   CDS             2..9691
FT                   /*tag=  b
FT   CDS             3..9692
FT                   /*tag=  c
XX
PN   USN7331212-N.
XX
PD   29-AUG-1989.
XX
PF   31-MAR-1989;   89US-0331212.
XX
PR   31-MAR-1989;   89US-0331212.
XX
PA   (USSH ) US DEPT. HEALTH AND HUMAN SERVICES.
XX
PI   Franchini G,  Wong-Staal F,  Gallo R;
XX
DR   WPI; 1989-339698/46.
DR   P-PSDB; AAP93283. AAP93284. AAP93285.
XX
PT   Complete human immunodeficiency type 2 proviral clone - used to generate
PT   animal model for function studies of HIV genes in vivo.
XX
PS   Disclosure; Fig. 5; 43pp; English.
XX
CC   HIV-2 SBL/ISY is a proviral clone of HIV-2. It contains a complete
CC   provirus including the cellular flanking sequences. The clone is
CC   infectious and cytopathic for some human T cell lines, induces syncytia
CC   and infects cell lines in vitro. The clone can be used to generate an
CC   animal model for functional studies of HIV genes in vivo, and to develop
CC   experimental approaches to prevent retroviral infection in man. It is
CC   constructed using DNA from neoplastic human cell line HUT78 freshly
CC   infected with the HIV-2 SBL6669 viral isolate. It can be translated in
CC   all 3 reading frames.
CC   (Note: Revised entry submitted to correct the patent number format of
CC   US Government-owned NTIS applications to prevent clashes with ongoing US
CC   granted patent numbers. For further information please visit the Derwent
CC   web site at www.derwent.com/dwpi/updates/ntis_us.html.)
XX
SQ   Sequence 9693 BP; 3315 A; 1947 C; 2412 G; 2019 T; 0 other;

  Query Match             51.4%;  Score 1356.4;  DB 10;  Length 9693;
  Best Local Similarity   72.8%;  Pred. No. 1.2e-301;
  Matches 1880;  Conservative    0;  Mismatches  626;  Indels   78;  Gaps    7;

Qy       2 CTAGAAGCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTGT 61
Db    6038 CTAGAAACATGCAATAACAAATGCTTTTGTAAGGGGTGCTGCTTCCATTGCCAGCTGTGT 6097

Qy      62 TTTCTTAAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCCG 121
Db    6098 TTTTTAAACAAGGGGCTCGGGATATGTTATGACCGAAAGGGCAGACGAAGAAGGAGTCCG 6157

Qy     122 AAAAAGGCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGGA 181
Db    6158 AAGAAAACTAAGGCTCATTCGTCTCCTGCATCAGACAAGTGAGTACAATGAGTGGTAAAA 6217

Qy     182 ATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATATG 241
Db    6218 TTCAGCTGCTTGTTGCCTTTCTGCTAACTAGTGCTTGCTTAATATATTGCACCAAATATG 6277

Qy     242 TCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCAA 301
Db    6278 TGACTGTTTTCTATGGAGTACCCGTGTGGAAAAATGCATCCATTCCCCTCTTTTGTGCAA 6337

Qy     302 CCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCAG 361
Db    6338 CTAAAAATAGAGATACTTGGGGAACCATACAGTGCTTGCCAGACAATGATGATTATCAAG 6397

Qy     362 AAGTGGCCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAGG 421
```

```
Db    6398 AGATACCTTTGAATGTAACAGAGGCTTTTGACGCATGGGATAATATAGTAACAGAACAAG 6457

Qy     422 CAATAGAGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCCC 481
Db    6458 CAGTAGAAGATGTCTGGAATCTATTTGAGACATCAATAAAACCATGTGTCAAATTAACGC 6517

Qy     482 CATTATGCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAAT 541
Db    6518 CTTTATGTGTAACAATGAACTGTAACGCAAGT---------------------------- 6549

Qy     542 CAATAACAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTCA 601
Db    6550 -----------ACAGAGAGCGCAGTTGCAACTACAAGCCCATCTGGACCTGATATGATAA 6598

Qy     602 ATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCAAATGA 661
Db    6599 ATGATACTGATCCATGCATACAATTGAACAATTGCTCAGGACTGAGGGAGGAAGACATGG 6658

Qy     662 TAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAATGAAA 721
Db    6659 TCGAGTGTCAGTTCAATATGACAGGACTAGAGTTAGATAAGAAAAAACAGTATAGTGAAA 6718

Qy     722 CTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGAT 781
Db    6719 CCTGGTACTCAAAAGATGTGGTTTGTGAATCAGATAACAGCACAG---ACCGAAAAAGAT 6775

Qy     782 GTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGGG 841
Db    6776 GTTACATGAACCATTGCAACACATCAGTCATCACAGAGTCATGTGACAAGCACTATTGGG 6835

Qy     842 ATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGACA 901
Db    6836 ATGCTATGAGATTTAGATACTGTGCACCACCGGGTTTTGTCTTGCTAAGGTGCAATGATA 6895

Qy     902 CAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGGA 961
Db    6896 CCAATTACTCAGGCTTTGAGCCCAATTGCTCTAAAGTAGTAGCTTCTACATGTACAAGAA 6955

Qy     962 TGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGAA 1021
Db    6956 TGATGGAAACGCAACCTTCTACTTGGCTTGGCTTTAATGGCACTAGGGCAGAAAATAGAA 7015

Qy    1022 CTTATATTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTATA 1081
Db    7016 CATATATCTATTGGCATGGTAGGGATAACAGAACTATTATCAGCTTAAACAAATATTATA 7075

Qy    1082 ATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATGT 1141
Db    7076 ATCTCACCATACTTTGTAGGAGACCAGAAAATAAAACAGTTGTACCAATAACACTCATGT 7135

Qy    1142 CTGGATTGGTTTTCCACTCAC---AACCAATCAATGATAGGCCAAAGCAGGCATGGTGTT 1198
Db    7136 CAGGCCGCAGATTTCACTCCCAGAAGATCATCAATAAAAAACCCAGGCAAGCATGGTGCC 7195

Qy    1199 GGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATC 1258
Db    7196 GGTTCAAAGGCGAGTGGAGGGAAGCCATGCAGGAGGTGAAACAAACCCTTGTAAAACATC 7255

Qy    1259 CCAGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGAGGAG--- 1315
Db    7256 CCAGGTATAAAGGAACCAATGACACAAATAAAATTAACTTTACAGCACCAGAAAAAGACT 7315

Qy    1316 GAGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAA 1375
Db    7316 CAGACCCAGAAGTAGCATATATGTGGACTAACTGCAGAGGAGAATTCCTCTATTGCAACA 7375

Qy    1376 TGAATTGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGGAAC 1435
Db    7376 TGACTTGGTTCCTTAATTGGGTAGAAAACAAGACGGGTCAACAGCATAAC---------- 7425

Qy    1436 AGCATAAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAG 1495
Db    7426 --------------TATGTGCCGTGCCATATAGAGCAAATAATTAATACCTGGCATAAGG 7471
```

```
Qy   1496 TAGGCAAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGA 1555
          |||| ||||||||| |||||||||||| || || |||||  |  | ||  | || |||||||
Db   7472 TAGGGAAAAATGTATATTTGCCTCCTAGGGAAGGAGAGTTGTCCTGCGAATCAACAGTGA 7531

Qy   1556 CCAGTCTCATAGCAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTG 1615
          ||||| |||| || ||||| |||      ||  |  |||| || ||||| ||| | ||||
Db   7532 CCAGTATCATTGCTAACATTGATGTTGATGGAGATAACCGGACAAATATTACCTTTAGTG 7591

Qy   1616 CAGAGGTGGCAGAACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTC 1675
          ||||||||||||||| || ||||||||||||| ||||||||||||||||  | || |
Db   7592 CAGAGGTGGCAGAACTATACCGATTGGAATTGGGGGATTATAAATTAGTAGAAGTAACAC 7651

Qy   1676 CAATTGGCTTGGCCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATA 1735
          |||||||||| ||||| |||| |  || || ||| |  ||| |    |    ||| |||
Db   7652 CAATTGGCTTCGCCCCTACAGCAGAAAAAAGATACTCCTCTGCT---CCAGGGAGACATA 7708

Qy   1736 AAAGAGGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGG 1795
          | ||||| ||  |||||||||||||| | ||||||||| | || |||||| |||||||||
Db   7709 AGAGAGGTGTGCTTGTGCTAGGGTTCCTAGGTTTTCTCACGACAGCAGGTGCTGCAATGG 7768

Qy   1796 GCGCGGCGTCGTTGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGC 1855
          | ||||||||  |||||||| | |||||||| || ||||||||   ||||||||||||||
Db   7769 GGGCGGCGTCTCTGACGCTGTCGGCTCAGTCTCGGACTTTATTCCGTGGGATAGTGCAGC 7828

Qy   1856 AACAGCAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCT 1915
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   7829 AACAGCAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAAATGTTGCGACTGACCGTCT 7888

Qy   1916 GGGGAACAAAGAACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGG 1975
          ||||||| || ||||||||| | || |||||||| || ||||||||| ||   |||||||
Db   7889 GGGGAACTAAAAACCTCCAGGCAAGAGTCACTGCTATTGAGAAGTACCTAGCAGACCAGG 7948

Qy   1976 CGCAGCTGAATGCTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGC 2035
          |||  || ||| | ||||||||||||||||||||||||||||||||||||||||||||||
Db   7949 CGCGACTAAATTCATGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGG 8008

Qy   2036 CAAATGCAAGTCTAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGG 2095
           |||||  |   |||||||  ||||||||||   ||| |||||||| ||||| |  ||
Db   8009 TAAATGACACCTTAACACCTGAGTGGAACAACATGACATGGCAAGAATGGGAACACAAAA 8068

Qy   2096 TTGACTTCTTGGAAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGA 2155
          |   |||| | || | |||||| |  |       |||  |||||||||| || ||||| |
Db   8069 TCCGCTTCCTAGAGGCAAATATCAGTGAGAGTTTAGAACAGGCACAAATCCAGCAAGAAA 8128

Qy   2156 AGAACATGTATGAATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACC 2215
          |||| ||||||||  | |||||| | ||||||||||||||| ||||||||||||||||||
Db   8129 AGAATATGTATGAGCTGCAAAAGCTAAATAGCTGGGATGTTTTTGGCAATTGGTTTGACT 8188

Qy   2216 TTGCTTCTTGGATAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGT 2275
          |  | || ||||| |||||||| ||||||||||    ||||| |||||| || ||
Db   8189 TAACCTCCTGGATCAAGTATATTCAATATGGAGTCATGATAGTAGTAGGAATAGTAGCTC 8248

Qy   2276 TAAGAATAGTGATCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAG 2335
          | |||||||| || ||| ||||||||||||||  ||   | ||  | || |||||||| |
Db   8249 TCAGAATAGTAATATATGTAGTACAAATGCTAAGTAGACTTAGAAAAGGCTATAGGCCTG 8308

Qy   2336 TGTTCTCTTCCCCACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGC 2395
             | |||   || |||  |||| | || |||| |||||||||  |||||  || || ||
Db   8309 ---TTTCTCTTCCCCCCGGTTATATTCAACAGATCCATATCCACAAGGACTGGGAACAGC 8365

Qy   2396 CAACCAGAGAAGGCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTT 2455
          ||  |||||||  | |||| ||||| ||| | | || ||| | ||| ||||||||||
Db   8366 CAGACAGAGAAGAAACAGAAGAAGACGTTGGGAACGACGTTGGAAGCAGATCCTGGCCTT 8425

Qy   2456 GGCAGATAGAATATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTAT 2515
          ||| |||||||||| ||||||||||||||| ||||| ||||||||||  | ||||
Db   8426 GGCCGATAGAATATATACATTTCCTGATCCGCCTGCTGATCCGCCTCTTGACCAGACTAT 8485

Qy   2516 TCAGCAACTGCAGAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGA 2575
          || || ||||||   ||| ||||| ||| | |||| ||||||||||| |     |
Db   8486 ACAACAGCTGCAGGGACTTACTATCCAGACTCTACCTGATCCTCCAACCACTCAGAGACT 8545

Qy   2576 GGCT 2579
          ||||
Db   8546 GGCT 8549
```

RESULT 15
AAC81937
ID AAC81937 standard; DNA; 9726 BP.
XX
AC AAC81937;
XX
DT 28-FEB-2001 (first entry)
XX
DE Packaging vector pROD(SD36/EM) DNA fragment.
XX
KW Encapsidation; transfer vector; nephrotropic; antiparkinsonian; anti-HIV;
KW cytostatic; gene therapy; transgenic; retroviral packaging;
KW gene delivery; Parkinson's disease; infectious diseases; cancer; ds.
XX
OS Synthetic.
XX
PN WO200040741-A2.
XX
PD 13-JUL-2000.
XX
PF 06-JAN-2000; 2000WO-US00390.
XX
PR 07-JAN-1999; 99US-0115247.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Arya SK;
XX
DR WPI; 2000-475836/41.
XX
PT New lentivirus transfer vector, functionally deleted for a splice donor
PT site and comprising a packaging signal and transgene operably linked to
PT a promoter, for improving encapsidation or transgene RNA and for gene
PT therapy -
XX
PS Example 6; Page 119-124; 143pp; English.
XX
CC This invention describes a novel transfer vector derived from a
CC lentivirus, functionally deleted for a splice donor site (SD), and
CC comprising a packaging signal and transgene operably linked to a
CC promoter. The products of the invention have nephrotropic,
CC antiparkinsonian, anti-HIV, and cytostatic activity and can be used for
CC gene therapy. Encapsidation of transgene RNA is improved using the new
CC retroviral packaging and transfer vectors. The new transfer and packaging
CC vectors are used as gene delivery agents and allows transfer of a
CC transgene into the genome of non-dividing cells. They can be used to
CC create a high-efficiency packaging cell line that provides greatly
CC enhanced packaging of foreign DNA. Individuals suffering from a
CC deficiency in alpha-galactosidase expression, such as Fabry disease can
CC be treated by delivering the vectors to cells in vitro or in vivo.
CC Parkinson's disease, infectious diseases, such as acquired
CC immunodeficiency syndrome and cancers can be treated with the vectors.
CC The non-infective packaging vectors can be used to detect wild-type HIV
CC in biological samples using southern or northern blot assays. The
CC packaging of the vector RNA is maximised, without an increase in the
CC packaging of the viral RNA. Deletion of sequences upstream and
CC downstream of the 5' SD region of the HIV-2 packaging vector results in
CC suppressed encapsidation of the packaging vector genomes without critical
CC loss of gene expression. Functional deletion of the SD site of the
CC transfer vector results in enhanced encapsidation of the transfer
CC vector's genome. HIV-2 packaging vector specifically and faithfully
CC packages its own optimally constructed transfer vector and gives better
CC quality and titre of vector than HIV-1.
XX
SQ Sequence 9726 BP; 3265 A; 1937 C; 2354 G; 1968 T; 202 other;

Query Match 51.0%; Score 1345.8; DB 21; Length 9726;
Best Local Similarity 72.3%; Pred. No. 3.2e-299;

```
 Matches 1911;  Conservative    0;  Mismatches  617;  Indels  115;  Gaps    8;

Qy      2 CTAGAAGCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTGT 61
          |||||| ||||| ||||| ||||||||||||||    || ||||||||||| ||| | |||
Db   5983 CTAGAAACATGCAATAACTCATGCTATTGTAAGCGATGCTGCTACCATTGTCAGATGTGT 6042

Qy     62 TTTCTTAAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCCG 121
          |||||| || || ||  | |||||||||||||| | |    | |   ||||||| |||||
Db   6043 TTTCTAAACAAGGGGCTCGGGATATGTTATGAACGAAAGGGCAGACGAAGAAGGACTCCA 6102

Qy    122 AAAAAGGCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGGA 181
          || ||  ||||| || || | ||| || || || |||||| ||||||           ||
Db   6103 AAGAAAACTAAGACTCATCCGTCTCCTACACCAGACAAGTGAGTATG---------ATGA 6153

Qy    182 ATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATATG 241
          |||||||||||||| ||||| ||  |    ||||  |    | ||||| || | |||||
Db   6154 ATCAGCTGCTTATTGCCATTTTATTAGCTAGTGCTTGCTTAGTATATTGCACCCAATATG 6213

Qy    242 TCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCAA 301
          | || || || ||||| |||||  | ||||  ||||| || ||||||||||||||||||||
Db   6214 TAACTGTTTTCTATGGCGTACCCACGTGGAAAAATGCAACCATTCCCCTCTTTTGTGCAA 6273

Qy    302 CCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCAG 361
          |||  |||||||||||||||||||||  |  |||||| | || || |||| |||||||   |
Db   6274 CCAGAAATAGGGATACTTGGGGAACCATACAGTGCTTGCCTGACAATGATGATTATCAGG 6333

Qy    362 AAGTGGCCCTTAATGTTACAGAAAGCTTTGATGCCT------------------------ 397
          || |  |  | ||||| |||||    |||||||| |.
Db   6334 AAATAACTTTGAATGTAACAGAGGCTTTTGATGCATGGNNTAGCGTGAGATCTTAGTGCA 6393

Qy    398 ------------------------------GGAATAATACAGTCACAGAACAGGCAATA 426
                                         ||||||||||||| |||||||| ||||||
Db   6394 TAGGTAGCGTGAGATGTTAGTGACTAAGATCGGAATAATACAGTAACAGAACAAGCAATA 6453

Qy    427 GAGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCCCCATTA 486
          || ||||| ||||| || || ||||| |||||||| || ||||| ||| || | || |||
Db   6454 GAAGATGTCTGGCATCTATTCGAGACATCAATAAAACCATGTGTCAAACTAACACCTTTA 6513

Qy    487 TGCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAATCAATA 546
          ||  |  | |||| |||||  |  |  |||| .  | | ||      ||||  ||||
Db   6514 TGTGTAGCAATGAAATGCAGCAGCACAGAGAGCAGCACAGGGAACAACACAACCTCAAAG 6573

Qy    547 ACAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTCAATGAG 606
          |  ||||        ||||| |||| ||| |  |||  | |  |   | | | ||||
Db   6574 AGCACAAGC------ACAACCACAACCACACCCACAGACCAGGAGCAAGAGATAAGTGAG 6627

Qy    607 ACTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCAAATGATAAGC 666
            || | | ||   |  |   || || ||| |||| |||| | | ||  ||| ||| |
Db   6628 GATACTCCATGCGCACGCGCAGACAACTGCTCAGGATTGGGAGAGGAAGAAACGATCAAT 6687

Qy    667 TGTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAATGAAACTTGG 726
          ||  | ||||| |||||||| ||| ||||||| |||||||| |||| |||||||| |||
Db   6688 TGCCAGTTCAATATGACAGGATTAGAAAGAGATAAGAAAAAACAGTATAATGAAACATGG 6747

Qy    727 TACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGATGTTAC 786
          |||||   |||| |||| |||||    |   |||| |||    ||| | |     ||||||
Db   6748 TACTCAAAAGATGTGGTTTGTGAGACAAATAATAGCAC---AAATCAGACCCAGTGTTAC 6804

Qy    787 ATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGGGATGCT 846
          |||||||| || ||||| || || |||  ||| || |||||||| || ||||||||||||
Db   6805 ATGAACCATTGCAACACATCAGTCATCACAGAATCATGTGACAAGCACTATTGGGATGCT 6864

Qy    847 ATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGACACAAAT 906
          || || ||||| || |||||||| || ||||||||  |  | |||||||||| || |||
Db   6865 ATAAGGTTTAGATACTGTGCACCACCGGGTTATGCCCTATTAAGATGTAATGATACCAAT 6924

Qy    907 TATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGGATGATG 966
          ||||||||||||   || || |||||||| ||.|| |  ||| ||||||| |||||||||
Db   6925 TATTCAGGCTTTGCACCCAACTGTTCTAAAGTAGTAGCTTCTACATGCACCAGGATGATG 6984

Qy    967 GAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGAACTTAT 1026
          || || || ||||| || |||||||||||||||| |||||||||| |||||||| |||
Db   6985 GAAACGCAAACTTCCACATGGTTTGGCTTTAATGGCACTAGAGCAGAGAATAGAACATAT 7044

Qy   1027 ATTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTATAATCTA 1086
          || || |||||||| || |||||||| ||||| || || ||||| || ||||||||||||
Db   7045 ATCTATTGGCATGGCAGAGATAATAGAACTATCATCAGCTTAAACAAATATTATAATCTC 7104

Qy   1087 ACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATGTCTGGA 1146
          |   || | ||||  || ||||| ||||||| |||  || | | |   ||||||| |||
Db   7105 AGTTTGCATTGTAAGAGGCCAGGGAATAAGATAGTGAAACAAATAATGCTTATGTCAGGA 7164

Qy   1147 TTGGTTTT------CCACTCACAACCAATCAATGATAGGCCAAAGCAGGCATGGTGTTGG 1200
             || ||      ||||| || || |||||| ||| || |  || |||||||| |||
Db   7165 CATGTGTTTCACTCCCACTACCAGCCGATCAATAAAAGACCCAGACAAGCATGGTGCTGG 7224

Qy   1201 TTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCC 1260
          ||   ||| |||||||| || || ||  | |||||||| | ||| |||  ||||||||
Db   7225 TTCAAAGGCAAATGGAAAGACGCCATGCAGGAGGTGAAGGAAACCCTTGCAAAACATCCC 7284

Qy   1261 AGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGAGGAGG---A 1317
          |||||||  |||||  || | ||    || || | ||  | || || |||  |||   |
Db   7285 AGGTATAGAGGAACCAATGACACAAGGAATATTAGCTTTGCAGCGCCAGGAAAAGGCTCA 7344

Qy   1318 GATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATG 1377
          || || |||||  | | |||||||||| || ||||||||||||||| ||||||| || |||
Db   7345 GACCCAGAAGTAGCATACATGTGGACTAACTGCAGAGGAGAGTTTCTCTACTGCAACATG 7404

Qy   1378 AATTGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGGAACAG 1437
          | |||||| || |||||| ||||  |||                              ||
Db   7405 ACTTGGTTCCTCAATTGGATAGAGAATA----------------------------AG 7434

Qy   1438 CATAAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTA 1497
             | | ||||| |  || || |||| |  |||||||| ||||| |||||||| |||
Db   7435 ACACACCGCAATTATGCACCGTGCCATATAAAGCAAATAATTAACACATGGCATAAGGTA 7494

Qy   1498 GGCAAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACC 1557
          || | |||||| |||||||||| || || || || ||  | || ||||| ||||  |||
Db   7495 GGGAGAAATGTATATTTGCCTCCCAGGGAAGGGGAGCTGTCCTGCAACTCAACAGTAACC 7554

Qy   1558 AGTCTCATAGCAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTGCA 1617
          ||  | || || ||||| || |||    |    ||:|| || || || ||| | ||||||
Db   7555 AGCATAATTGCTAACATTGACTGGCAAAACAATAATCAGACAAACATTACCTTTAGTGCA 7614

Qy   1618 GAGGTGGCAGAACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTCCA 1677
          |||||||||||||| ||  |||||| |||||||||||||||| ||||| || || |||
Db   7615 GAGGTGGCAGAACTATACAGATTGGAGTTGGGAGATTATAAATTGGTAGAAATAACACCA 7674

Qy   1678 ATTGGCTTGGCCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAA 1737
          |||||||| || || ||| | |  || || ||| |.  ||| |    |  ||| ||| |
Db   7675 ATTGGCTTCGCACCTACAAAAGAAAAAAGATACTCCTCTGCT---CACGGGAGACATACA 7731

Qy   1738 AGAGGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGGGC 1797
          ||||| || || |||||||||||||||||||||||||||||| ||||||||||||||||||
Db   7732 AGAGGTGTGTTCGTGCTAGGGTTCTTGGGTTTTCTCGCAACAGCAGGTTCTGCAATGGGC 7791

Qy   1798 GCGGCGTCGTTGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGCAA 1857
          |||||||  |||| || | |||||||||| |||||| |||| ||||||||||||||||||
Db   7792 GCGGCGTCCCTGACCGTGTCGGCTCAGTCCCGGACTTTACTGGCCGGGATAGTGCAGCAA 7851

Qy   1858 CAGCAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCTGG 1917
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   7852 CAGCAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAACTGTTGCGACTGACCGTCTGG 7911

Qy   1918 GGAACAAAGAACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGGCG 1977
          ||||| || |||||||| | || |||||||| || |||||||| || ||||||||||||
Db   7912 GGAACGAAAAACCTCCAGGCAAGAGTCACTGCTATAGAGAAGTACCTACAGGACCAGGCG 7971

Qy   1978 CAGCTGAATGCTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGCCA 2037
          | ||| ||| | |||||||||||| ||||||||||||||||||||||||||||||||||
Db   7972 CGGCTAAATTCATGGGGATGTGCGGTTAGACAAGTCTGCCACACTACTGTACCATGGGTT 8031

Qy   2038 AATGCAAGTCTAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTT 2097
          ||||      || |||| | ||| |||| ||| ||||| || |||||   | | ||
Db   8032 AATGATTCCTTAGCACCTGACTGGGACAATATGACGTGGCAGGAATGGGAAAAACAAGTC 8091
```

```
Qy   2098 GACTTCTTGGAAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAG 2157
          || | |||| | |||||| |          ||||  ||||||||||||| ||||||||
Db   8092 CGCTACCTGGAGGCAAATATCAGTAAAAGTTTAGAACAGGCACAAATTCAGCAAGAGAAA 8151

Qy   2158 AACATGTATGAATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTT 2217
          || |||||||||| ||||||| || |||||||||||| | ||||||||||||||||||| |
Db   8152 AATATGTATGAACTACAAAAATTAAATAGCTGGGATATTTTTGGCAATTGGTTTGACTTA 8211

Qy   2218 GCTTCTTGGATAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTA 2277
          | || ||| | |||||||| |||||||||||   |||| | |||| |||||||   |||
Db   8212 ACCTCCTGGGTCAAGTATATTCAATATGGAGTGCTTATAATAGTAGCAGTAATAGCTTTA 8271

Qy   2278 AGAATAGTGATCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTG 2337
          |||||||||||| ||| |||||||||||| ||  || | | ||  |||| |||||||| ||
Db   8272 AGAATAGTGATATATGTAGTACAAATGTTAAGTAGGCTTAGAAAGGGCTATAGGCCTGTT 8331

Qy   2338 TTCTCTTCCCCACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCA 2397
          |||||||||||| |||  |||| |||| |||| ||||||||  |||||| || || ||||
Db   8332 TTCTCTTCCCCCCCCGGTTATATCCAACAGATCCATATCCACAAGGACCGGGGACAGCCA 8391

Qy   2398 ACCAGAGAAGGCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGG 2457
          |||  ||||  | ||||  |||||||||     ||||||  |||| | |||||| |||
Db   8392 GCCAACGAAGAAACAGAAGAAGACGGTGGAAGCAACGGTGGAGACAGATACTGGCCCTGG 8451

Qy   2458 CAGATAGAATATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTC 2517
          | ||||| |||||| |||||||||||||||||| ||||| ||||||||||  | |||| |
Db   8452 CCGATAGCATATATACATTTCCTGATCCGCCAGCTGATTCGCCTCTTGACCAGACTATAC 8511

Qy   2518 AGCAACTGCAGAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGG 2577
          |||| ||||||   ||| ||||| ||    | || || |||||||  ||   |||||
Db   8512 AGCATCTGCAGGGACTTACTATCCAGGAGCTTCCTGACCCTCCAACTCATCTACCAGACA 8571

Qy   2578 CTC 2580
          ||
Db   8572 GTC 8574


Search completed: November 16, 2002, 08:04:34
Job time : 860.611 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 16, 2002, 06:34:38 ; Search time 61.2201 Seconds
(without alignments)
13229.850 Million cell updates/sec

Title: M33262_COPY_6695_9335
Perfect score: 2641
Sequence: 1 TCTAGAAGCATGCTATAACA.....ACTGACCTACCTACAATATG 2641

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

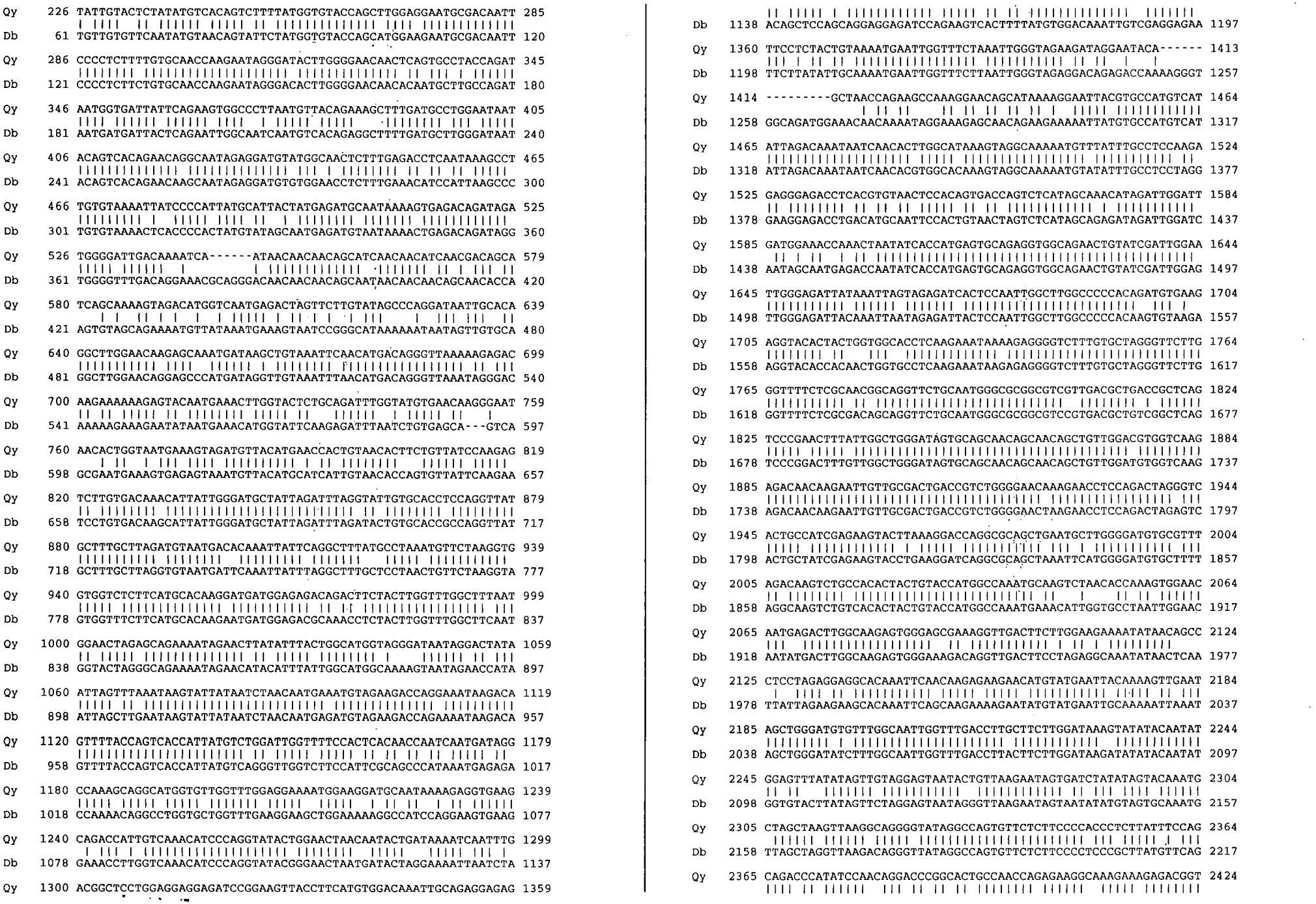
	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1745.4	66.1	2658	1	US-08-254-358-2	Sequence 2, Appli
	2	1745.4	66.1	2658	1	US-08-475-391-2	Sequence 2, Appli
	3	1745.4	66.1	2658	2	US-08-709-609-2	Sequence 2, Appli
	4	1745.4	66.1	2658	5	PCT-US95-07178-2	Sequence 2, Appli
	5	1356.4	51.4	9633	6	5223423-1	Patent No. 5223423
	6	1346.4	51.0	10163	2	US-08-659-251-1	Sequence 1, Appli
	7	1346.4	51.0	10163	4	US-09-256-490-1	Sequence 1, Appli
	8	1346.4	51.0	10163	5	PCT-US96-11445-1	Sequence 1, Appli
	9	1266.6	48.0	2574	2	US-08-659-251-13	Sequence 13, Appl
	10	1266.6	48.0	2574	4	US-09-256-490-13	Sequence 13, Appl
	11	1266.6	48.0	2574	5	PCT-US96-11445-13	Sequence 13, Appl
	12	251.6	9.5	4527	2	US-08-944-449-8	Sequence 8, Appli
	13	251.6	9.5	4527	4	US-09-353-362-8	Sequence 8, Appli
	14	242	9.2	2571	1	US-08-254-358-3	Sequence 3, Appli
	15	242	9.2	2571	1	US-08-475-391-3	Sequence 3, Appli
	16	242	9.2	2571	2	US-08-709-609-3	Sequence 3, Appli
	17	242	9.2	2571	5	PCT-US95-07178-3	Sequence 3, Appli
c	18	242	9.2	6474	3	US-08-651-472-66	Sequence 66, Appl
c	19	242	9.2	6474	4	US-08-358-928-66	Sequence 66, Appl
	20	242	9.2	6926	3	US-08-651-472-69	Sequence 69, Appl
	21	242	9.2	6926	4	US-08-358-928-69	Sequence 69, Appl
	22	242	9.2	9739	1	US-08-022-835-1	Sequence 1, Appli
	23	242	9.2	9739	1	US-08-388-809-1	Sequence 1, Appli
	24	242	9.2	9739	2	US-08-647-714-1	Sequence 1, Appli
	25	239.2	9.1	9737	2	US-08-944-449-7	Sequence 7, Appli
	26	239.2	9.1	9737	4	US-09-353-362-7	Sequence 7, Appli
	27	237.2	9.0	9746	1	US-08-022-835-3	Sequence 3, Appli
	28	237.2	9.0	9746	1	US-08-388-809-3	Sequence 3, Appli
	29	237.2	9.0	9746	2	US-08-647-714-3	Sequence 3, Appli
	30	234.4	8.9	3563	3	US-08-463-210-6	Sequence 6, Appli
	31	232.6	8.8	2571	2	US-07-916-098A-1	Sequence 1, Appli
	32	232.2	8.8	3807	1	US-08-022-835-5	Sequence 5, Appli
	33	232.2	8.8	3807	1	US-08-388-809-5	Sequence 5, Appli
	34	232.2	8.8	3807	2	US-08-647-714-5	Sequence 5, Appli
	35	231	8.7	2570	2	US-08-448-603A-31	Sequence 31, Appl
	36	231	8.7	2570	3	US-09-134-075-31	Sequence 31, Appl
	37	231	8.7	2570	4	US-09-492-739-31	Sequence 31, Appl
	38	226.2	8.6	8932	4	US-09-124-900-1	Sequence 1, Appli
	39	226.2	8.6	8933	3	US-08-463-210-4	Sequence 4, Appli
	40	226.2	8.6	8933	4	US-09-620-958A-3	Sequence 3, Appli
	41	226.2	8.6	8933	4	US-09-620-958A-4	Sequence 4, Appli
	42	226.2	8.6	8933	4	US-09-620-958A-9	Sequence 9, Appli
	43	224.6	8.5	2552	2	US-08-448-603A-27	Sequence 27, Appl
	44	224.6	8.5	2552	3	US-09-134-075-27	Sequence 27, Appl
	45	224.6	8.5	2552	4	US-09-492-739-27	Sequence 27, Appl

ALIGNMENTS

RESULT 1
US-08-254-358-2
; Sequence 2, Application US/08254358
; Patent No. 5658785
; GENERAL INFORMATION:
; APPLICANT: Johnson, Philip R.
; TITLE OF INVENTION: Adeno-Associated Virus Materials and
; TITLE OF INVENTION: Methods
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 S. Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/254,358
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5658785and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31975
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2658 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-254-358-2

Query Match 66.1%; Score 1745.4; DB 1; Length 2658;
Best Local Similarity 82.2%; Pred. No. 0;
Matches 2052; Conservative 0; Mismatches 421; Indels 24; Gaps 3;

```
Qy    166 ATGGGATGTCTTGGGAATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATC 225
          ||||||||||||||||||||||||||||||||  |||||||  |||||||  | | |||
Db      1 ATGGGATGTCTTGGGAATCAGCTGCTTATCGCGCTCTTGCTAGTAAGTGTTTTAGAGATT 60
```



```
Db   2218 CAGATCCCTATCCACAAGGGCCAGGAACCGCCAACCAAAGAAGGAGAAGAAGGAGACGGT 2277

Qy   2425 GGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAATATATTCATTTCCTGATC 2484
          |||||  | |||||||  ||| || ||||||||||||||||||||||||||||||||||||
Db   2278 GGAGACAGAGGTGGCAGCAGATCTTGGCCTTGGCAGATAGAATATATTCATTTCCTGATC 2337

Qy   2485 CGCCAACTGATACGCCTCTTGACTTGGCTATTCAGCAACTGCAGAACCTTGCTATCGAGA 2544
          |||||  |||||||||||||||||||||||||||||| ||||||    | |||||  |||
Db   2338 CGCCAGTTGATACGCCTCTTGACTTGGCTATTCAGCAGCTGCAGGGATTGGCTATTGAGG 2397

Qy   2545 GTATACCAGATCCTCCAACCAATACTCCAGAGGCTCTCTGCGACCCTACAGAGGATTCGA 2604
             |||||||||||||||||  | |||||||| |||||  ||||  | || ||  | ||
Db   2398 AGCTACCAGATCCTCCAACCAGTGCTCCAGAGCCTCTCAACGACGTTGCAAAGAGTCCGT 2457

Qy   2605 GAAGTCCTCAGGACTGAACTGACCTACCTACAATATG 2641
          ||||||  ||||  |  ||||  |  ||||||||||||||||
Db   2458 GAAGTCATCAGAATTGAAATAGCCTACCTACAATATG 2494


RESULT 2
US-08-475-391-2
; Sequence 2, Application US/08475391
; Patent No. 5786211
;  GENERAL INFORMATION:
;    APPLICANT:  Johnson, Philip R.
;    TITLE OF INVENTION:  Adeno-Associated Virus Materials and
;    TITLE OF INVENTION:  Methods
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 S. Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/475,391
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/254,358
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5786211and, Greta E.
;      REGISTRATION NUMBER:  35,302
;      REFERENCE/DOCKET NUMBER:  31975
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 474-6300
;      TELEFAX:  (312) 474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2658 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-475-391-2

  Query Match             66.1%;  Score 1745.4;  DB 1;  Length 2658;
  Best Local Similarity   82.2%;  Pred. No. 0;
  Matches 2052;  Conservative    0;  Mismatches  421;  Indels   24;  Gaps    3;

Qy    166 ATGGGATGTCTTGGGAATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATC 225
          |||||||||||||||||||||||||||||||  |||||||  |||||||  |  | |||
Db      1 ATGGGATGTCTTGGGAATCAGCTGCTTATCGCGCTCTTGCTAGTAAGTGTTTTAGAGATT 60

Qy    226 TATTGTACTCTATATGTCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATT 285
          |||| ||  |||||| ||||| || ||||||||||||| |||| ||||||||||||||
Db     61 TGTTGTGTTCAATATGTAACAGTATTCTATGGTGTACCAGCATGGAAGAATGCGACAATT 120

Qy    286 CCCCTCTTTTGTGCAACCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGAT 345
          |||||||| |||||||||||||||||| |||||||||||||| || ||| | ||||||
Db    121 CCCCTCTTCTGTGCAACCAAGAATAGGGACACTTGGGGAACAACACAATGCTTGCCAGAT 180

Qy    346 AATGGTGATTATTCAGAAGTGGCCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAAT 405
          |||| |||||| |||||| ||||  | ||||| |||||    |||||||| ||| ||||||
Db    181 AATGATGATTACTCAGAATTGGCAATCAATGTCACAGAGGCTTTTGATGCTTGGGATAAT 240

Qy    406 ACAGTCACAGAACAGGCAATAGAGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCT 465
          |||||||||||||| |||||||||||||| ||| | |||||||| || || || ||||||
Db    241 ACAGTCACAGAACAAGCAATAGAGGATGTGTGGAACCTCTTTGAAACATCCATTAAGCCC 300

Qy    466 TGTGTAAAATTATCCCCATTATGCATTACTATGAGATGCAATAAAAGTGAGACAGATAGA 525
          |||||||||  ||||| |||| ||  | |||||||| |||||||  ||||||||||||||
Db    301 TGTGTAAAACTCACCCCACTATGTATAGCAATGAGATGTAATAAAACTGAGACAGATAGG 360

Qy    526 TGGGGATTGACAAAATCA------ATAACAACAACAGCATCAACAACATCAACGACAGCA 579
          |||||  ||||||  |         | ||||||||||||||  |||||||  || | ||| ||
Db    361 TGGGGTTTGACAGGAAACGCAGGGACAACAACAACAGCAATAACAACAACAGCAACACCA 420

Qy    580 TCAGCAAAAGTAGACATGGTCAATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACA 639
            | |  || | |  |  | ||||| | || | |  | |||    |  ||| |||  ||
Db    421 AGTGTAGCAGAAAATGTTATAAATGAAAGTAATCCGGGCATAAAAAATAATAGTTGTGCA 480

Qy    640 GGCTTGGAACAAGAGCAAATGATAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGAC 699
          ||||||||||| ||||  |||||| | |||||||| ||||||||||||||||  || |||
Db    481 GGCTTGGAACAGGAGCCCATGATAGGTTGTAAATTTAACATGACAGGGTTAAATAGGGAC 540

Qy    700 AAGAAAAAAGAGTACAATGAAACTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAAT 759
          || || ||||| || |||||||| ||||| ||   ||||||  | ||||| ||  |
Db    541 AAAAAGAAAGAATATAATGAAACATGGTATTCAAGAGATTTAATCTGTGAGCA---GTCA 597

Qy    760 AACACTGGTAATGAAAGTAGATGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAG 819
            | ||  | ||| |||| |||||||||| | || ||||||||   |||||| |||||
Db    598 GCGAATGAAAGTGAGAGTAAATGTTACATGCATCATTGTAACACCAGTGTTATTCAAGAA 657

Qy    820 TCTTGTGACAAACATTATTGGGATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTAT 879
          || |||||||| |||||||||||||||||||||||||||||| || |||||||| |||||||
Db    658 TCCTGTGACAAGCATTATTGGGATGCTATTAGATTTAGATACTGTGCACCGCCAGGTTAT 717

Qy    880 GCTTTGCTTAGATGTAATGACACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTG 939
          |||||||||| ||||||||  |||||||| |||||||   ||||| ||||||||||
Db    718 GCTTTGCTTAGGTGTAATGATTCAAATTATTTAGGCTTTGCTCCTAACTGTTCTAAGGTA 777

Qy    940 GTGGTCTCTTCATGCACAAGGATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAAT 999
          ||||| |||||||||||||| ||||||||||| || ||.|||||||||||||||||| |||
Db    778 GTGGTTTCTTCATGCACAAGAATGATGGAGACGCAAACCTCTACTTGGTTTGGCTTCAAT 837

Qy   1000 GGAACTAGAGCAGAAAATAGAACTTATATTTACTGGCATGGTAGGGATAATAGGACTATA 1059
          || ||||| |||||||||||||| || ||||| |||||||| |    |||||| || |||
Db    838 GGTACTAGGGCAGAAAATAGAACATACATTTATTGGCATGGCAAAAGTAATAGAACCATA 897

Qy   1060 ATTAGTTTAAATAAGTATTATAATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACA 1119
          ||||| || |||||||||||||||||||||||| |||||||||||||| ||||||||||
Db    898 ATTAGCTTGAATAAGTATTATAATCTAACAATGAGATGTAGAAGACCAGAAAATAAGACA 957

Qy   1120 GTTTTACCAGTCACCATTATGTCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGG 1179
          |||||||||||||||||||||| || ||||| ||||| || || || || ||||| ||
Db    958 GTTTTACCAGTCACCATTATGTCAGGGTTGGTCTTCCATTCGCAGCCCATAAATGAGAGA 1017

Qy   1180 CCAAAGCAGGCATGGTGTTGGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAG 1239
          ||||| ||||| ||||| ||||||| |||||  |||||  | || ||  | || ||||||
Db   1018 CCAAAACAGGCCTGGTGCTGGTTTGAAGGAAGCTGGAAAAAGGCCATCCAGGAAGTGAAG 1077

Qy   1240 CAGACCATTGTCAAACATCCCAGGTATACTGGAACTAACAATACTGATAAAATCAATTTG 1299
            | ||| | |||||||||||||||||||| ||||||||  |||||   ||||| ||| |
Db   1078 GAAACCTTGGTCAAACATCCCAGGTATACGGGAACTAATGATACTAGGAAAATTAATCTA 1137

Qy   1300 ACGGCTCCTGGAGGAGGAGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAG 1359
          || ||||| | |||||||||||| |||||  || || ||||||||||||||  |||||||
```

```
Db   1138 ACAGCTCCAGCAGGAGGAGATCCAGAAGTCACTTTTATGTGGACAAATTGTCGAGGAGAA 1197

Qy   1360 TTCCTCTACTGTAAAATGAATTGGTTTCTAAATTGGGTAGAAGATAGGAATACA------ 1413
          ||| | || || |||||||||||||||| |||||||||| || ||  |   |
Db   1198 TTCTTATATTGCAAAATGAATTGGTTTCTTAATTGGGTAGAGGACAGAGACCAAAAGGGT 1257

Qy   1414 ---------GCTAACCAGAAGCCAAAGGAACAGCATAAAAGGAATTACGTGCCATGTCAT 1464
                    | || ||    || || || || || |  ||||| ||||||||||||
Db   1258 GGCAGATGGAAACAACAAAATAGGAAAGAGCAACAGAAGAAAAATTATGTGCCATGTCAT 1317

Qy   1465 ATTAGACAAATAATCAACACTTGGCATAAAGTAGGCAAAAATGTTTATTTGCCTCCAAGA 1524
          |||||||||||||||||| ||||| |||||||||||||||| |||||||||| ||
Db   1318 ATTAGACAAATAATCAACACGTGGCACAAAGTAGGCAAAAATGTATATTTGCCTCCTAGG 1377

Qy   1525 GAGGGAGACCTCACGTGTAACTCCACAGTGACCAGTCTCATAGCAAACATAGATTGGATT 1584
          || |||||||| || || || ||||| || || |||||||||||| | ||||||||||
Db   1378 GAAGGAGACCTGACATGCAATTCCACTGTAACTAGTCTCATAGCAGAGATAGATTGGATC 1437

Qy   1585 GATGGAAACCAAACTAATATCACCATGAGTGCAGAGGTGGCAGAACTGTATCGATTGGAA 1644
            || | ||  | || ||||||||||||||||||||||||||||||||||||||||||
Db   1438 AATAGCAATGAGACCAATATCACCATGAGTGCAGAGGTGGCAGAACTGTATCGATTGGAG 1497

Qy   1645 TTGGGAGATTATAAATTAGTAGAGATCACTCCAATTGGCTTGGCCCCCACAGATGTGAAG 1704
          |||||||||| |||||| ||||||| |||||||||||||||||||||||  ||| |
Db   1498 TTGGGAGATTACAAATTAATAGAGATTACTCCAATTGGCTTGGCCCCCACAAGTGTAAGA 1557

Qy   1705 AGGTACACTACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCTTTGTGCTAGGGTTCTTG 1764
          |||||||| ||   |||  |||||||||||| ||||||||||||||||||||||||||||
Db   1558 AGGTACACCACAACTGGTGCCTCAAGAAATAAGAGAGGGGTCTTTGTGCTAGGGTTCTTG 1617

Qy   1765 GGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGTTGACGCTGACCGCTCAG 1824
          |||||||||| || |||||||||||||||||||||||||  ||||||||  | ||||||
Db   1618 GGTTTTCTCGCGACAGCAGGTTCTGCAATGGGCGCGGCGTCCGTGACGCTGTCGGCTCAG 1677

Qy   1825 TCCCGAACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGCTGTTGGACGTGGTCAAG 1884
          ||||| ||||| |||||||||||||||||||||||||||||||||||||| |||||||||
Db   1678 TCCCGGACTTTGTTGGCTGGGATAGTGCAGCAACAGCAACAGCTGTTGGATGTGGTCAAG 1737

Qy   1885 AGACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACAAAGAACCTCCAGACTAGGGTC 1944
          |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||| |||
Db   1738 AGACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACTAAGAACCTCCAGACTAGAGTC 1797

Qy   1945 ACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATGCTTGGGGATGTGCGTTT 2004
          ||||| |||||||||||| | ||||| |||||||||| ||| | |||||||||||| |||
Db   1798 ACTGCTATCGAGAAGTACCTGAAGGATCAGGCGCAGCTAAATTCATGGGGATGTGCTTTT 1857

Qy   2005 AGACAAGTCTGCCACACTACTGTACCATGGCCAAATGCAAGTCTAACACCAAAGTGGAAC 2064
          || |||||||| |||||||||||||||||||||||| ||   |    || || ||||||
Db   1858 AGGCAAGTCTGTCACACTACTGTACCATGGCCAAATGAAACATTGGTGCCTAATTGGAAC 1917

Qy   2065 AATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGGAAGAAAATATAACAGCC 2124
          |||  |||||||||||||||||  || |||||||||| | || | |||||||||
Db   1918 AATATGACTTGGCAAGAGTGGGAAAGACAGGTTGACTTCCTAGAGGCAAATATAACTCAA 1977

Qy   2125 CTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATGAATTACAAAAGTTGAAT 2184
            |  |||| || |||||||||| ||||| ||||| |||||||||| ||||| || |||
Db   1978 TTATTAGAAGAAGCACAAATTCAGCAAGAAAAGAATATGTATGAATTGCAAAAATTAAAT 2037

Qy   2185 AGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGATAAAGTATATACAATAT 2244
          ||||||||| | |||||||||||||||||||| |||||||||||  ||||||||||||
Db   2038 AGCTGGGATATCTTTGGCAATTGGTTTGACCTTACTTCTTGGATAAGATATATACAATAT 2097

Qy   2245 GGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGATCTATATAGTACAAATG 2304
          || ||   ||||||| |||||||||||  |||||||||| || ||| |||| ||||||
Db   2098 GGTGTACTTATAGTTCTAGGAGTAATAGGGTTAAGAATAGTAATATATGTAGTGCAAATG 2157

Qy   2305 CTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTCTCTTCCCCACCCTCTTATTTCCAG 2364
            |||||| |||||| ||||| |||||||||||||||||||||| ||| ||||| | |||
Db   2158 TTAGCTAGGTTAAGACAGGGTTATAGGCCAGTGTTCTCTTCCCCTCCCGCTTATGTTCAG 2217

Qy   2365 CAGACCCATATCCAACAGGACCCGGCACTGCCAACCAGAGAAGGCAAAGAAAGAGACGGT 2424
          |||| || ||||||  ||| || || || |||||||| ||||||  ||||| ||||||||
Db   2218 CAGATCCCTATCCACAAGGGCCAGGAACCGCCAACCAAAGAAGGAGAAGAAGGAGACGGT 2277

Qy   2425 GGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAATATATTCATTTCCTGATC 2484
          |||||  | ||||||| ||| || ||||||||||||||||||||||||||||||||||||
Db   2278 GGAGACAGAGGTGGCAGCAGATCTTGGCCTTGGCAGATAGAATATATTCATTTCCTGATC 2337

Qy   2485 CGCCAACTGATACGCCTCTTGACTTGGCTATTCAGCAACTGCAGAACCTTGCTATCGAGA 2544
          |||||  |||||||||||||||||||||||||||||| ||||||    | ||||| |||
Db   2338 CGCCAGTTGATACGCCTCTTGACTTGGCTATTCAGCAGCTGCAGGGATTGGCTATTGAGG 2397

Qy   2545 GTATACCAGATCCTCCAACCAATACTCCAGAGGCTCTCTGCGACCCTACAGAGGATTCGA 2604
             |||||||||||||||||| | |||||||| ||||  ||||  | || ||  | ||
Db   2398 AGCTACCAGATCCTCCAACCAGTGCTCCAGAGCCTCTCAACGACGTTGCAAAGAGTCCGT 2457

Qy   2605 GAAGTCCTCAGGACTGAACTGACCTACCTACAATATG 2641
          ||||||  |||| | |||| |  ||||||||||||||
Db   2458 GAAGTCATCAGAATTGAAATAGCCTACCTACAATATG 2494
```

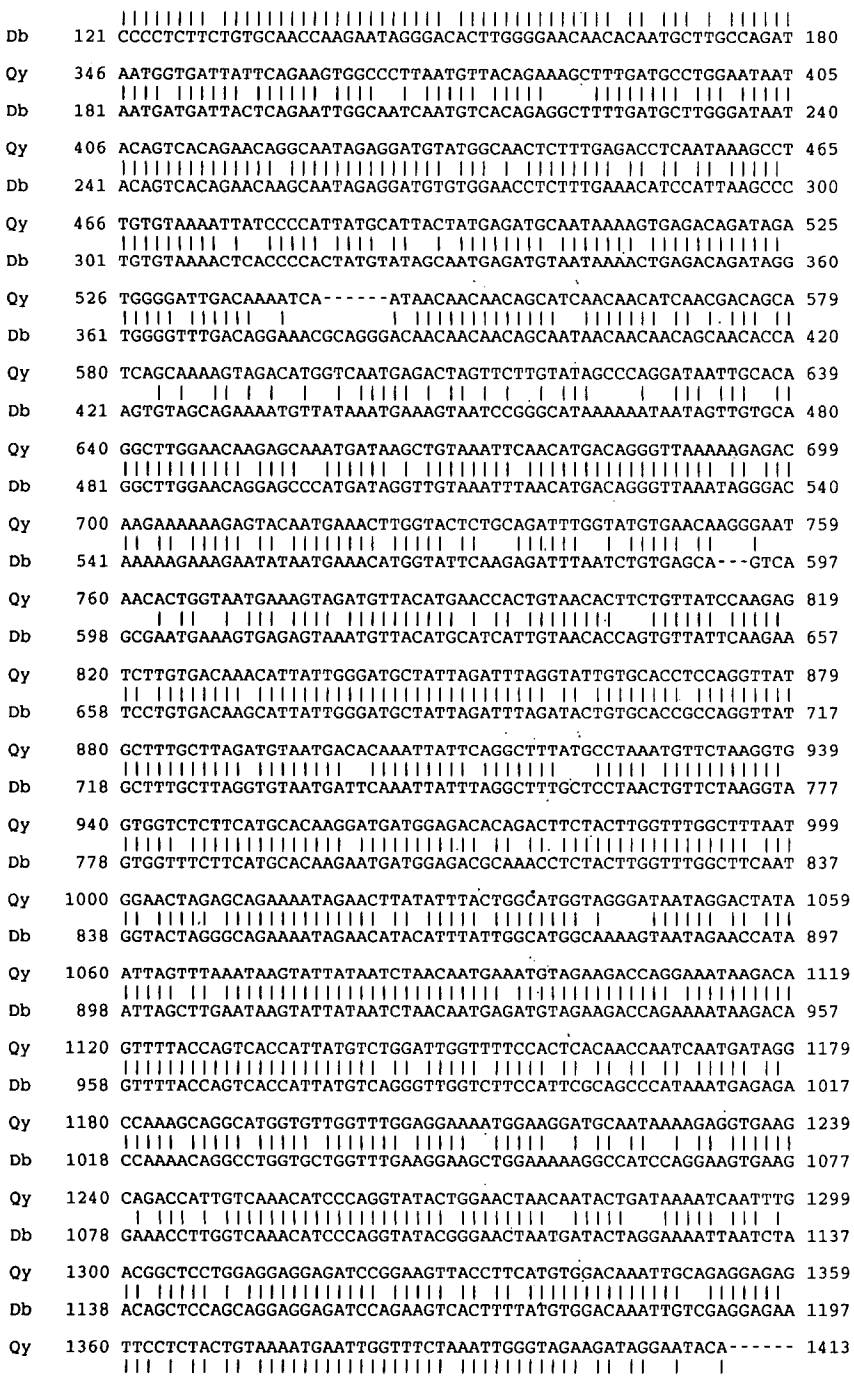
RESULT 3
US-08-709-609-2
; Sequence 2, Application US/08709609
; Patent No. 5858775
; GENERAL INFORMATION:
; APPLICANT: Johnson, Philip R.
; TITLE OF INVENTION: Adeno-Associated Virus Materials and
; TITLE OF INVENTION: Methods
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 S. Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/709,609
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5858775and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31975
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2658 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-709-609-2

Query Match 66.1%; Score 1745.4; DB 2; Length 2658;
Best Local Similarity 82.2%; Pred. No. 0;
Matches 2052; Conservative 0; Mismatches 421; Indels 24; Gaps 3;

```
Qy    166 ATGGGATGTCTTGGGAATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATC 225
          |||||||||||||||||||||||||||||||  |||||||  |||||||  |  | |||
Db      1 ATGGGATGTCTTGGGAATCAGCTGCTTATCGCGCTCTTGCTAGTAAGTGTTTTAGAGATT 60

Qy    226 TATTGTACTCTATATGTCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATT 285
          | ||||  || |||||| ||||| || ||||||||||||||| |||| ||||||||||||
Db     61 TGTTGTGTTCAATATGTAACAGTATTCTATGGTGTACCAGCATGGAAGAATGCGACAATT 120

Qy    286 CCCCTCTTTTGTGCAACCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGAT 345
```


Db 121 CCCCTCTTCTGTGCAACCAAGAATAGGGACACTTGGGGAACAACACAATGCTTGCCAGAT 180
Qy 346 AATGGTGATTATTCAGAAGTGGCCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAAT 405
Db 181 AATGATGATTACTCAGAATTGGCAATCAATGTCACAGAGGCTTTTGATGCTTGGGATAAT 240
Qy 406 ACAGTCACAGAACAGGCAATAGAGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCT 465
Db 241 ACAGTCACAGAACAAGCAATAGAGGATGTGTGGAACCTCTTTGAAACATCCATTAAGCCC 300
Qy 466 TGTGTAAAATTATCCCCATTATGCATTACTATGAGATGCAATAAAAGTGAGACAGATAGA 525
Db 301 TGTGTAAAACTCACCCCACTATGTATAGCAATGAGATGTAATAAAACTGAGACAGATAGG 360
Qy 526 TGGGGATTGACAAAATCA------ATAACAACAACAGCATCAACAACATCAACGACAGCA 579
Db 361 TGGGGTTTGACAGGAAACGCAGGGACAACAACAACAGCAATAACAACAACAGCAACACCA 420
Qy 580 TCAGCAAAAGTAGACATGGTCAATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACA 639
Db 421 AGTGTAGCAGAAAATGTTATAAATGAAAGTAATCCGGGCATAAAAAATAATAGTTGTGCA 480
Qy 640 GGCTTGGAACAAGAGCAAATGATAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGAC 699
Db 481 GGCTTGGAACAGGAGCCCATGATAGGTTGTAAATTTAACATGACAGGGTTAAATAGGGAC 540
Qy 700 AAGAAAAAAGAGTACAATGAAACTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAAT 759
Db 541 AAAAAGAAAGAATATAATGAAACATGGTATTCAAGAGATTTAATCTGTGAGCA---GTCA 597
Qy 760 AACACTGGTAATGAAAGTAGATGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAG 819
Db 598 GCGAATGAAAGTGAGAGTAAATGTTACATGCATCATTGTAACACCAGTGTTATTCAAGAA 657
Qy 820 TCTTGTGACAAACATTATTGGGATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTAT 879
Db 658 TCCTGTGACAAGCATTATTGGGATGCTATTAGATTTAGATACTGTGCACCGCCAGGTTAT 717
Qy 880 GCTTTGCTTAGATGTAATGACACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTG 939
Db 718 GCTTTGCTTAGGTGTAATGATTCAAATTATTTAGGCTTTGCTCCTAACTGTTCTAAGGTA 777
Qy 940 GTGGTCTCTTCATGCACAAGGATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAAT 999
Db 778 GTGGTTTCTTCATGCACAAGAATGATGGAGACGCAAACCTCTACTTGGTTTGGCTTCAAT 837
Qy 1000 GGAACTAGAGCAGAAAATAGAACTTATATTTACTGGCATGGTAGGGATAATAGGACTATA 1059
Db 838 GGTACTAGGGCAGAAAATAGAACATACATTTATTGGCATGGCAAAAGTAATAGAACCATA 897
Qy 1060 ATTAGTTTAAATAAGTATTATAATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACA 1119
Db 898 ATTAGCTTGAATAAGTATTATAATCTAACAATGAGATGTAGAAGACCAGAAAATAAGACA 957
Qy 1120 GTTTTACCAGTCACCATTATGTCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGG 1179
Db 958 GTTTTACCAGTCACCATTATGTCAGGGTTGGTCTTCCATTCGCAGCCCATAAATGAGAGA 1017
Qy 1180 CCAAAGCAGGCATGGTGTTGGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAG 1239
Db 1018 CCAAAACAGGCCTGGTGCTGGTTTGAAGGAAGCTGGAAAAAGGCCATCCAGGAAGTGAAG 1077
Qy 1240 CAGACCATTGTCAAACATCCCAGGTATACTGGAACTAACAATACTGATAAAATCAATTTG 1299
Db 1078 GAAACCTTGGTCAAACATCCCAGGTATACGGGAACTAATGATACTAGGAAAATTAATCTA 1137
Qy 1300 ACGGCTCCTGGAGGAGGAGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAG 1359
Db 1138 ACAGCTCCAGCAGGAGGAGATCCAGAAGTCACTTTTATGTGGACAAATTGTCGAGGAGAA 1197
Qy 1360 TTCCTCTACTGTAAAATGAATTGGTTTCTAAATTGGGTAGAAGATAGGAATACA------ 1413
Db 1198 TTCTTATATTGCAAAATGAATTGGTTTCTTAATTGGGTAGAGGACAGAGACCAAAAGGGT 1257
Qy 1414 ---------GCTAACCAGAAGCCAAAGGAACAGCATAAAAGGAATTACGTGCCATGTCAT 1464
Db 1258 GGCAGATGGAAACAACAAAATAGGAAAGAGCAACAGAAGAAAAATTATGTGCCATGTCAT 1317
Qy 1465 ATTAGACAAATAATCAACACTTGGCATAAAGTAGGCAAAAATGTTTATTTGCCTCCAAGA 1524
Db 1318 ATTAGACAAATAATCAACACGTGGCACAAAGTAGGCAAAAATGTATATTTGCCTCCTAGG 1377
Qy 1525 GAGGGAGACCTCACGTGTAACTCCACAGTGACCAGTCTCATAGCAAACATAGATTGGATT 1584
Db 1378 GAAGGAGACCTGACATGCAATTCCACTGTAACTAGTCTCATAGCAGAGATAGATTGGATC 1437
Qy 1585 GATGGAAACCAAACTAATATCACCATGAGTGCAGAGGTGGCAGAACTGTATCGATTGGAA 1644
Db 1438 AATAGCAATGAGACCAATATCACCATGAGTGCAGAGGTGGCAGAACTGTATCGATTGGAG 1497
Qy 1645 TTGGGAGATTATAAATTAGTAGAGATCACTCCAATTGGCTTGGCCCCCACAGATGTGAAG 1704
Db 1498 TTGGGAGATTACAAATTAATAGAGATTACTCCAATTGGCTTGGCCCCCACAAGTGTAAGA 1557
Qy 1705 AGGTACACTACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCTTTGTGCTAGGGTTCTTG 1764
Db 1558 AGGTACACCACAACTGGTGCCTCAAGAAATAAGAGAGGGGTCTTTGTGCTAGGGTTCTTG 1617
Qy 1765 GGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGTTGACGCTGACCGCTCAG 1824
Db 1618 GGTTTTCTCGCGACAGCAGGTTCTGCAATGGGCGCGGCGTCCGTGACGCTGTCGGCTCAG 1677
Qy 1825 TCCCGAACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGCTGTTGGACGTGGTCAAG 1884
Db 1678 TCCCGGACTTTGTTGGCTGGGATAGTGCAGCAACAGCAACAGCTGTTGGATGTGGTCAAG 1737
Qy 1885 AGACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACAAAGAACCTCCAGACTAGGGTC 1944
Db 1738 AGACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACTAAGAACCTCCAGACTAGAGTC 1797
Qy 1945 ACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATGCTTGGGGATGTGCGTTT 2004
Db 1798 ACTGCTATCGAGAAGTACCTGAAGGATCAGGCGCAGCTAAATTCATGGGGATGTGCTTTT 1857
Qy 2005 AGACAAGTCTGCCACACTACTGTACCATGGCCAAATGCAAGTCTAACACCAAAGTGGAAC 2064
Db 1858 AGGCAAGTCTGTCACACTACTGTACCATGGCCAAATGAAACATTGGTGCCTAATTGGAAC 1917
Qy 2065 AATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGGAAGAAAATATAACAGCC 2124
Db 1918 AATATGACTTGGCAAGAGTGGGAAAGACAGGTTGACTTCCTAGAGGCAAATATAACTCAA 1977
Qy 2125 CTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATGAATTACAAAAGTTGAAT 2184
Db 1978 TTATTAGAAGAAGCACAAATTCAGCAAGAAAAGAATATGTATGAATTGCAAAAATTAAAT 2037
Qy 2185 AGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGATAAAGTATATACAATAT 2244
Db 2038 AGCTGGGATATCTTTGGCAATTGGTTTGACCTTACTTCTTGGATAAGATATATACAATAT 2097
Qy 2245 GGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGATCTATATAGTACAAATG 2304
Db 2098 GGTGTACTTATAGTTCTAGGAGTAATAGGGTTAAGAATAGTAATATATGTAGTGCAAATG 2157
Qy 2305 CTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTCTCTTCCCCACCCTCTTATTTCCAG 2364
Db 2158 TTAGCTAGGTTAAGACAGGGTTATAGGCCAGTGTTCTCTTCCCCTCCCGCTTATGTTCAG 2217
Qy 2365 CAGACCCATATCCAACAGGACCCGGCACTGCCAACCAGAGAAGGCAAAGAAAGAGACGGT 2424
Db 2218 CAGATCCCTATCCACAAGGGCCAGGAACCGCCAACCAAAGAAGGAGAAGAAGGAGACGGT 2277
Qy 2425 GGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAATATATTCATTTCCTGATC 2484
Db 2278 GGAGACAGAGGTGGCAGCAGATCTTGGCCTTGGCAGATAGAATATATTCATTTCCTGATC 2337

```
Qy   2485 CGCCAACTGATACGCCTCTTGACTTGGCTATTCAGCAACTGCAGAACCTTGCTATCGAGA 2544
          |||||  |||||||||||||||||||||||||||||||| ||||||    | ||||| |||
Db   2338 CGCCAGTTGATACGCCTCTTGACTTGGCTATTCAGCAGCTGCAGGGATTGGCTATTGAGG 2397

Qy   2545 GTATACCAGATCCTCCAACCAATACTCCAGAGGCTCTCTGCGACCCTACAGAGGATTCGA 2604
               |||||||||||||||| | |||||||| |||||  ||||  | || ||  | ||
Db   2398 AGCTACCAGATCCTCCAACCAGTGCTCCAGAGCCTCTCAACGACGTTGCAAAGAGTCCGT 2457

Qy   2605 GAAGTCCTCAGGACTGAACTGACCTACCTACAATATG 2641
          |||||| |||| | |||| |  ||||||||||||||
Db   2458 GAAGTCATCAGAATTGAAATAGCCTACCTACAATATG 2494


RESULT 4
PCT-US95-07178-2
; Sequence 2, Application PC/TUS9507178
;  GENERAL INFORMATION:
;    APPLICANT:  Johnson, Philip R.
;    TITLE OF INVENTION:  Adeno-Associated Virus Materials and
;    TITLE OF INVENTION:  Methods
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 S. Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/07178
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Noland, Greta E.
;      REGISTRATION NUMBER:  35,302
;      REFERENCE/DOCKET NUMBER:  31975
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 474-6300
;      TELEFAX:  (312) 474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2658 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
PCT-US95-07178-2

  Query Match             66.1%;  Score 1745.4;  DB 5;  Length 2658;
  Best Local Similarity   82.2%;  Pred. No. 0;
  Matches 2052;  Conservative    0;  Mismatches  421;  Indels   24;  Gaps    3;

Qy    166 ATGGGATGTCTTGGGAATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATC 225
          |||||||||||||||||||||||||||||||||||||  |||||||  |||||||  |  | |||
Db      1 ATGGGATGTCTTGGGAATCAGCTGCTTATCGCGCTCTTGCTAGTAAGTGTTTTAGAGATT 60

Qy    226 TATTGTACTCTATATGTCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATT 285
          | ||||  || ||||||  ||||| || ||||||||||||||| |||| ||||||||||||
Db     61 TGTTGTGTTCAATATGTAACAGTATTCTATGGTGTACCAGCATGGAAGAATGCGACAATT 120

Qy    286 CCCCTCTTTTGTGCAACCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGAT 345
          |||||||| |||||||||||||||||||| ||||||||||||| || ||| | ||||||
Db    121 CCCCTCTTCTGTGCAACCAAGAATAGGGACACTTGGGGAACAACACAATGCTTGCCAGAT 180

Qy    346 AATGGTGATTATTCAGAAGTGGCCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAAT 405
          |||| |||||| |||||| ||||  | ||||| |||||    |||||||| ||| |||||

Db    181 AATGATGATTACTCAGAATTGGCAATCAATGTCACAGAGGCTTTTGATGCTTGGGATAAT 240

Qy    406 ACAGTCACAGAACAGGCAATAGAGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCT 465
          ||||||||||||| |||||||||||||| ||| | |||||||| || || || |||||
Db    241 ACAGTCACAGAACAAGCAATAGAGGATGTGTGGAACCTCTTTGAAACATCCATTAAGCCC 300

Qy    466 TGTGTAAAATTATCCCCATTATGCATTACTATGAGATGCAATAAAAGTGAGACAGATAGA 525
          ||||||||| |  ||||| |||| ||  | |||||||| ||||||| ||||||||||||
Db    301 TGTGTAAAACTCACCCCACTATGTATAGCAATGAGATGTAATAAAACTGAGACAGATAGG 360

Qy    526 TGGGGATTGACAAAATCA------ATAACAACAACAGCATCAACAACATCAACGACAGCA 579
          ||||| ||||||  |          | ||||||||||||  ||||||| || | ||| ||
Db    361 TGGGGTTTGACAGGAAACGCAGGGACAACAACAACAGCAATAACAACAACAGCAACACCA 420

Qy    580 TCAGCAAAAGTAGACATGGTCAATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACA 639
          | | || | |  | ||||| | || | |  | |||    |  ||| ||| ||
Db    421 AGTGTAGCAGAAAATGTTATAAATGAAAGTAATCCGGGCATAAAAAATAATAGTTGTGCA 480

Qy    640 GGCTTGGAACAAGAGCAAATGATAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGAC 699
          |||||||||| |||| |||||| | |||||||| |||||||||||||||||| || |||
Db    481 GGCTTGGAACAGGAGCCCATGATAGGTTGTAAATTTAACATGACAGGGTTAAATAGGGAC 540

Qy    700 AAGAAAAAAGAGTACAATGAAACTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAAT 759
          || || ||||| || |||||||| ||||| ||   ||||||  | ||||| ||   |
Db    541 AAAAAGAAAGAATATAATGAAACATGGTATTCAAGAGATTTAATCTGTGAGCA---GTCA 597

Qy    760 AACACTGGTAATGAAAGTAGATGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAG 819
          | ||  | ||| |||| ||||||||| | || |||||||   |||||| |||||
Db    598 GCGAATGAAAGTGAGAGTAAATGTTACATGCATCATTGTAACACCAGTGTTATTCAAGAA 657

Qy    820 TCTTGTGACAAACATTATTGGGATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTAT 879
          || |||||||| |||||||||||||||||||||||| || |||||||| |||||||||
Db    658 TCCTGTGACAAGCATTATTGGGATGCTATTAGATTTAGATACTGTGCACCGCCAGGTTAT 717

Qy    880 GCTTTGCTTAGATGTAATGACACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTG 939
          |||||||||| ||||||||  |||||||| |||||||   ||||| ||||||||||
Db    718 GCTTTGCTTAGGTGTAATGATTCAAATTATTTAGGCTTTGCTCCTAACTGTTCTAAGGTA 777

Qy    940 GTGGTCTCTTCATGCACAAGGATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAAT 999
          ||||| |||||||||||||| |||||||||| || || |||||||||||||||||| |||
Db    778 GTGGTTTCTTCATGCACAAGAATGATGGAGACGCAAACCTCTACTTGGTTTGGCTTCAAT 837

Qy   1000 GGAACTAGAGCAGAAAATAGAACTTATATTTACTGGCATGGTAGGGATAATAGGACTATA 1059
          || ||||| |||||||||||||| || ||||| |||||||| |    |||||| || |||
Db    838 GGTACTAGGGCAGAAAATAGAACATACATTTATTGGCATGGCAAAAGTAATAGAACCATA 897

Qy   1060 ATTAGTTTAAATAAGTATTATAATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACA 1119
          ||||| || |||||||||||||||||||||||| |||||||||||||| ||||||||||
Db    898 ATTAGCTTGAATAAGTATTATAATCTAACAATGAGATGTAGAAGACCAGAAAATAAGACA 957

Qy   1120 GTTTTACCAGTCACCATTATGTCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGG 1179
          |||||||||||||||||||||||| || ||||| |||| || || || || ||||| ||
Db    958 GTTTTACCAGTCACCATTATGTCAGGGTTGGTCTTCCATTCGCAGCCCATAAATGAGAGA 1017

Qy   1180 CCAAAGCAGGCATGGTGTTGGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAG 1239
          ||||| ||||| ||||| ||||||| ||||| |||||  | || ||  | || ||||||
Db    1018 CCAAAACAGGCCTGGTGCTGGTTTGAAGGAAGCTGGAAAAAGGCCATCCAGGAAGTGAAG 1077

Qy   1240 CAGACCATTGTCAAACATCCCAGGTATACTGGAACTAACAATACTGATAAAATCAATTTG 1299
           | ||| | |||||||||||||||||||| ||||||||  |||||   ||||| ||| |
Db    1078 GAAACCTTGGTCAAACATCCCAGGTATACGGGAACTAATGATACTAGGAAAATTAATCTA 1137

Qy   1300 ACGGCTCCTGGAGGAGGAGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAG 1359
          || ||||| | |||||||||||| ||||| || || ||||||||||||||  |||||||
Db    1138 ACAGCTCCAGCAGGAGGAGATCCAGAAGTCACTTTTATGTGGACAAATTGTCGAGGAGAA 1197

Qy   1360 TTCCTCTACTGTAAAATGAATTGGTTTCTAAATTGGGTAGAAGATAGGAATACA------ 1413
          ||| | || || |||||||||||||||||| |||||||||| || ||  |   |
Db    1198 TTCTTATATTGCAAAATGAATTGGTTTCTTAATTGGGTAGAGGACAGAGACCAAAAGGGT 1257

Qy   1414 ---------GCTAACCAGAAGCCAAAGGAACAGCATAAAAGGAATTACGTGCCATGTCAT 1464
                    | || ||    || || || || || | ||||| ||||||||||||
Db    1258 GGCAGATGGAAACAACAAAATAGGAAAGAGCAACAGAAGAAAAATTATGTGCCATGTCAT 1317
```

```
Qy   1465 ATTAGACAAATAATCAACACTTGGCATAAAGTAGGCAAAAATGTTTATTTGCCTCCAAGA 1524
          |||||||||||||||||||| ||||| |||||||||||||||| |||||||||| ||
Db   1318 ATTAGACAAATAATCAACACGTGGCACAAAGTAGGCAAAAATGTATATTTGCCTCCTAGG 1377

Qy   1525 GAGGGAGACCTCACGTGTAACTCCACAGTGACCAGTCTCATAGCAAACATAGATTGGATT 1584
          || |||||||| || || || ||||| || || |||||||||||| | ||||||||||
Db   1378 GAAGGAGACCTGACATGCAATTCCACTGTAACTAGTCTCATAGCAGAGATAGATTGGATC 1437

Qy   1585 GATGGAAACCAAACTAATATCACCATGAGTGCAGAGGTGGCAGAACTGTATCGATTGGAA 1644
           || | ||  | || ||||||||||||||||||||||||||||||||||||||||||||
Db   1438 AATAGCAATGAGACCAATATCACCATGAGTGCAGAGGTGGCAGAACTGTATCGATTGGAG 1497

Qy   1645 TTGGGAGATTATAAATTAGTAGAGATCACTCCAATTGGCTTGGCCCCCACAGATGTGAAG 1704
          |||||||||| |||||| ||||||| |||||||||||||||||||||||||  ||| |
Db   1498 TTGGGAGATTACAAATTAATAGAGATTACTCCAATTGGCTTGGCCCCCACAAGTGTAAGA 1557

Qy   1705 AGGTACACTACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCTTTGTGCTAGGGTTCTTG 1764
          |||||||| ||   |||  |||||||||||| |||||||||||||||||||||||||||
Db   1558 AGGTACACCACAACTGGTGCCTCAAGAAATAAGAGAGGGGTCTTTGTGCTAGGGTTCTTG 1617

Qy   1765 GGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGTTGACGCTGACCGCTCAG 1824
          |||||||||| || |||||||||||||||||||||||||||||  ||||||| | ||||||
Db   1618 GGTTTTCTCGCGACAGCAGGTTCTGCAATGGGCGCGGCGTCCGTGACGCTGTCGGCTCAG 1677

Qy   1825 TCCCGAACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGCTGTTGGACGTGGTCAAG 1884
          ||||| |||||  ||||||||||||||||||||||||||||||||||||||| |||||||||
Db   1678 TCCCGGACTTTGTTGGCTGGGATAGTGCAGCAACAGCAACAGCTGTTGGATGTGGTCAAG 1737

Qy   1885 AGACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACAAAGAACCTCCAGACTAGGGTC 1944
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||| |||
Db   1738 AGACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACTAAGAACCTCCAGACTAGAGTC 1797

Qy   1945 ACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATGCTTGGGGATGTGCGTTT 2004
          ||||| |||||||||||| | ||||| |||||||||||| ||| | ||||||||||| |||
Db   1798 ACTGCTATCGAGAAGTACCTGAAGGATCAGGCGCAGCTAAATTCATGGGGATGTGCTTTT 1857

Qy   2005 AGACAAGTCTGCCACACTACTGTACCATGGCCAAATGCAAGTCTAACACCAAAGTGGAAC 2064
          || |||||||| |||||||||||||||||||||||||| ||    |   || || ||||||
Db   1858 AGGCAAGTCTGTCACACTACTGTACCATGGCCAAATGAAACATTGGTGCCTAATTGGAAC 1917

Qy   2065 AATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGGAAGAAAATATAACAGCC 2124
          |||  |||||||||||||||||  || |||||||||| | || | |||||||||
Db   1918 AATATGACTTGGCAAGAGTGGGAAAGACAGGTTGACTTCCTAGAGGCAAATATAACTCAA 1977

Qy   2125 CTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATGAATTACAAAAGTTGAAT 2184
          |  |||| || |||||||||| |||| ||||| |||||||||| ||||| || |||
Db   1978 TTATTAGAAGAAGCACAAATTCAGCAAGAAAAGAATATGTATGAATTGCAAAAATTAAAT 2037

Qy   2185 AGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGATAAAGTATATACAATAT 2244
          |||||||| | ||||||||||||||||||||| |||||||||||  ||||||||||||
Db   2038 AGCTGGGATATCTTTGGCAATTGGTTTGACCTTACTTCTTGGATAAGATATATACAATAT 2097

Qy   2245 GGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGATCTATATAGTACAAATG 2304
          || ||   ||||||| |||||||||||  |||||||||||| || ||| |||| ||||||
Db   2098 GGTGTACTTATAGTTCTAGGAGTAATAGGGTTAAGAATAGTAATATATGTAGTGCAAATG 2157

Qy   2305 CTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTCTCTTCCCCACCCTCTTATTTCCAG 2364
          ||||||| |||||| |||| |||||||||||||||||||||||||| ||| ||||| | |||
Db   2158 TTAGCTAGGTTAAGACAGGGTTATAGGCCAGTGTTCTCTTCCCCTCCCGCTTATGTTCAG 2217

Qy   2365 CAGACCCATATCCAACAGGACCCGGCACTGCCAACCAGAGAAGGCAAAGAAAGAGACGGT 2424
          |||| || ||||||  ||| || || || |||||||| ||||||  ||||| ||||||||
Db   2218 CAGATCCCTATCCACAAGGGCCAGGAACCGCCAACCAAAGAAGGAGAAGAAGGAGACGGT 2277

Qy   2425 GGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAATATATTCATTTCCTGATC 2484
          |||||  | ||||||| ||| || ||||||||||||||||||||||||||||||||||||
Db   2278 GGAGACAGAGGTGGCAGCAGATCTTGGCCTTGGCAGATAGAATATATTCATTTCCTGATC 2337

Qy   2485 CGCCAACTGATACGCCTCTTGACTTGGCTATTCAGCAACTGCAGAACCTTGCTATCGAGA 2544
          |||||  |||||||||||||||||||||||||||||||||||| ||||||    | ||||| |||
Db   2338 CGCCAGTTGATACGCCTCTTGACTTGGCTATTCAGCAGCTGCAGGGATTGGCTATTGAGG 2397

Qy   2545 GTATACCAGATCCTCCAACCAATACTCCAGAGGCTCTCTGCGACCCTACAGAGGATTCGA 2604
          ||||||||||||||||||| | |||||||| |||||  ||||  | || ||  | ||
Db   2398 AGCTACCAGATCCTCCAACCAGTGCTCCAGAGCCTCTCAACGACGTTGCAAAGAGTCCGT 2457

Qy   2605 GAAGTCCTCAGGACTGAACTGACCTACCTACAATATG 2641
          |||||| |||| | |||| |  ||||||||||||||
Db   2458 GAAGTCATCAGAATTGAAATAGCCTACCTACAATATG 2494


RESULT 5
5223423-1
;Patent No. 5223423
;    APPLICANT: FRANCHINI, GENOVEFFA;WONG-STAAL, FLOSSIE;
;GALLO, ROBERT
;    TITLE OF INVENTION: CHARACTERIZATION OF REPLICATION COMPETENT
;HUMAN IMMUNODEFICIENCY TYPE 2 PROVIRAL CLONE HIV-2 SBL/ISY
;    NUMBER OF SEQUENCES: 4
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/331,212
;      FILING DATE: 03-31-1989
;SEQ ID NO:1:
;      LENGTH: 9633
5223423-1

  Query Match            51.4%;  Score 1356.4;  DB 6;  Length 9633;
  Best Local Similarity  72.8%;  Pred. No. 0;
  Matches 1880;  Conservative    0;  Mismatches  626;  Indels   78;  Gaps    7;

Qy      2 CTAGAAGCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTGT 61
          |||||| ||||| |||||| ||||| ||||||   ||| |||| ||||||||||  | |||
Db   5978 CTAGAAACATGCAATAACAAATGCTTTTGTAAGGGGTGCTGCTTCCATTGCCAGCTGTGT 6037

Qy     62 TTTCTTAAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCCG 121
          ||| | || || ||  | |||||||||||||| | |    | |   ||||||| | |||||
Db   6038 TTTTTAAACAAGGGGCTCGGGATATGTTATGACCGAAAGGGCAGACGAAGAAGGAGTCCG 6097

Qy    122 AAAAAGGCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGGA 181
          || ||  |||||||| || | ||| |||||||| |||||| ||||      ||  |   |
Db   6098 AAGAAAACTAAGGCTCATTCGTCTCCTGCATCAGACAAGTGAGTACAATGAGTGGTAAAA 6157

Qy    182 ATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATATG 241
          ||||||||||  | ||| |  ||||    ||||  |    || ||||| ||   |||||
Db   6158 TTCAGCTGCTTGTTGCCTTTCTGCTAACTAGTGCTTGCTTAATATATTGCACCAAATATG 6217

Qy    242 TCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCAA 301
          | || || || ||||| ||||| |  ||||  |||||  | ||||||||||||||||||
Db   6218 TGACTGTTTTCTATGGAGTACCCGTGTGGAAAAATGCATCCATTCCCCTCTTTTGTGCAA 6277

Qy    302 CCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCAG 361
          | || ||||| ||||||||||||||  |  |||||| | ||||| |||| |||||||  ||
Db   6278 CTAAAAATAGAGATACTTGGGGAACCATACAGTGCTTGCCAGACAATGATGATTATCAAG 6337

Qy    362 AAGTGGCCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAGG 421
          |  |  |  | ||||| |||||    ||||| || ||| |||||| ||| |||||||| |
Db   6338 AGATACCTTTGAATGTAACAGAGGCTTTTGACGCATGGGATAATATAGTAACAGAACAAG 6397

Qy    422 CAATAGAGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCCC 481
          || |||| ||||| ||| | || |||||||| |||||||| || |||| |||||| | |
Db   6398 CAGTAGAAGATGTCTGGAATCTATTTGAGACATCAATAAAACCATGTGTCAAATTAACGC 6457

Qy    482 CATTATGCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAAT 541
          | |||||  | || ||||  || ||   ||||
Db   6458 CTTTATGTGTAACAATGAACTGTAACGCAAGT---------------------------- 6489

Qy    542 CAATAACAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTCA 601
                |||||       ||    |||| |||   || |   |   || ||| | |
Db   6490 -----------ACAGAGAGCGCAGTTGCAACTACAAGCCCATCTGGACCTGATATGATAA 6538

Qy    602 ATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCAAATGA 661
          |||| |||  | | || |||     | | |||||| |||| ||    | ||  ||||
Db   6539 ATGATACTGATCCATGCATACAATTGAACAATTGCTCAGGACTGAGGGAGGAAGACATGG 6598

Qy    662 TAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAATGAAA 721
```

```
Db   6599 TCGAGTGTCAGTTCAATATGACAGGACTAGAGTTAGATAAGAAAAAACAGTATAGTGAAA 6658

Qy    722 CTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGAT 781

Db   6659 CCTGGTACTCAAAAGATGTGGTTTGTGAATCAGATAACAGCACAG---ACCGAAAAAGAT 6715

Qy    782 GTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGGG 841

Db   6716 GTTACATGAACCATTGCAACACATCAGTCATCACAGAGTCATGTGACAAGCACTATTGGG 6775

Qy    842 ATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGACA 901

Db   6776 ATGCTATGAGATTTAGATACTGTGCACCACCGGGTTTTGTCTTGCTAAGGTGCAATGATA 6835

Qy    902 CAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGGA 961

Db   6836 CCAATTACTCAGGCTTTGAGCCCAATTGCTCTAAAGTAGTAGCTTCTACATGTACAAGAA 6895

Qy    962 TGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGAA 1021

Db   6896 TGATGGAAACGCAACCTTCTACTTGGCTTGGCTTTAATGGCACTAGGGCAGAAAATAGAA 6955

Qy   1022 CTTATATTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTATA 1081

Db   6956 CATATATCTATTGGCATGGTAGGGATAACAGAACTATTATCAGCTTAAACAAATATTATA 7015

Qy   1082 ATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATGT 1141

Db   7016 ATCTCACCATACTTTGTAGGAGACCAGAAAATAAAACAGTTGTACCAATAACACTCATGT 7075

Qy   1142 CTGGATTGGTTTTCCACTCAC---AACCAATCAATGATAGGCCAAAGCAGGCATGGTGTT 1198

Db   7076 CAGGCCGCAGATTTCACTCCCAGAAGATCATCAATAAAAAACCCAGGCAAGCATGGTGCC 7135

Qy   1199 GGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATC 1258

Db   7136 GGTTCAAAGGCGAGTGGAGGGAAGCCATGCAGGAGGTGAAACAAACCCTTGTAAAACATC 7195

Qy   1259 CCAGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGAGGAG--- 1315

Db   7196 CCAGGTATAAAGGAACCAATGACACAAATAAAATTAACTTTACAGCACCAGAAAAAGACT 7255

Qy   1316 GAGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAA 1375

Db   7256 CAGACCCAGAAGTAGCATATATGTGGACTAACTGCAGAGGAGAATTCCTCTATTGCAACA 7315

Qy   1376 TGAATTGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGGAAC 1435

Db   7316 TGACTTGGTTCCTTAATTGGGTAGAAAACAAGACGGGTCAACAGCATAAC---------- 7365

Qy   1436 AGCATAAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAG 1495

Db   7366 --------------TATGTGCCGTGCCATATAGAGCAAATAATTAATACCTGGCATAAGG 7411

Qy   1496 TAGGCAAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGA 1555

Db   7412 TAGGGAAAAATGTATATTTGCCTCCTAGGGAAGGAGAGTTGTCCTGCGAATCAACAGTGA 7471

Qy   1556 CCAGTCTCATAGCAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTG 1615

Db   7472 CCAGTATCATTGCTAACATTGATGTTGATGGAGATAACCGGACAAATATTACCTTTAGTG 7531

Qy   1616 CAGAGGTGGCAGAACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTC 1675

Db   7532 CAGAGGTGGCAGAACTATACCGATTGGAATTGGGGGATTATAAATTAGTAGAAGTAACAC 7591

Qy   1676 CAATTGGCTTGGCCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATA 1735

Db   7592 CAATTGGCTTCGCCCCTACAGCAGAAAAAAGATACTCCTCTGCT---CCAGGGAGACATA 7648

Qy   1736 AAAGAGGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGG 1795

Db   7649 AGAGAGGTGTGCTTGTGCTAGGGTTCCTAGGTTTTCTCACGACAGCAGGTGCTGCAATGG 7708

Qy   1796 GCGCGGCGTCGTTGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGC 1855

Db   7709 GGGCGGCGTCTCTGACGCTGTCGGCTCAGTCTCGGACTTTATTCCGTGGGATAGTGCAGC 7768

Qy   1856 AACAGCAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCT 1915

Db   7769 AACAGCAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAAATGTTGCGACTGACCGTCT 7828

Qy   1916 GGGGAACAAAGAACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGG 1975

Db   7829 GGGGAACTAAAAACCTCCAGGCAAGAGTCACTGCTATTGAGAAGTACCTAGCAGACCAGG 7888

Qy   1976 CGCAGCTGAATGCTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGC 2035

Db   7889 CGCGACTAAATTCATGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGG 7948

Qy   2036 CAAATGCAAGTCTAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGG 2095

Db   7949 TAAATGACACCTTAACACCTGAGTGGAACAACATGACATGGCAAGAATGGGAACACAAAA 8008

Qy   2096 TTGACTTCTTGGAAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGA 2155

Db   8009 TCCGCTTCCTAGAGGCAAATATCAGTGAGAGTTTAGAACAGGCACAAATCCAGCAAGAAA 8068

Qy   2156 AGAACATGTATGAATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACC 2215

Db   8069 AGAATATGTATGAGCTGCAAAAGCTAAATAGCTGGGATGTTTTTGGCAATTGGTTTGACT 8128

Qy   2216 TTGCTTCTTGGATAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGT 2275

Db   8129 TAACCTCCTGGATCAAGTATATTCAATATGGAGTCATGATAGTAGTAGGAATAGTAGCTC 8188

Qy   2276 TAAGAATAGTGATCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAG 2335

Db   8189 TCAGAATAGTAATATATGTAGTACAAATGCTAAGTAGACTTAGAAAAGGCTATAGGCCTG 8248

Qy   2336 TGTTCTCTTCCCCACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGC 2395

Db   8249 ---TTTCTCTTCCCCCCGGTTATATTCAACAGATCCATATCCACAAGGACTGGGAACAGC 8305

Qy   2396 CAACCAGAGAAGGCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTT 2455

Db   8306 CAGACAGAGAAGAAACAGAAGAAGACGTTGGGAACGACGTTGGAAGCAGATCCTGGCCTT 8365

Qy   2456 GGCAGATAGAATATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTAT 2515

Db   8366 GGCCGATAGAATATATACATTTCCTGATCCGCCTGCTGATCCGCCTCTTGACCAGACTAT 8425

Qy   2516 TCAGCAACTGCAGAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGA 2575

Db   8426 ACAACAGCTGCAGGGACTTACTATCCAGACTCTACCTGATCCTCCAACCACTCAGAGACT 8485

Qy   2576 GGCT 2579

Db   8486 GGCT 8489
```

RESULT 6
US-08-659-251-1
; Sequence 1, Application US/08659251
; Patent No. 5883081
; GENERAL INFORMATION:
; APPLICANT: Kraus, Guenter
; APPLICANT: Wong-Staal, Flossie
; APPLICANT: Talbott, Randy
; APPLICANT: Poeschla, Eric
; TITLE OF INVENTION: Isolation of No. 5883081el HIV-2 Proviruses
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor

```
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/659,251
;      FILING DATE:  No. 5883081 yet assigned
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/001,441
;      FILING DATE:  26-JUL-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Garrett-Wackowski, Eugenia
;      REGISTRATION NUMBER:  37,330
;      REFERENCE/DOCKET NUMBER:  02307E-056410US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 576-0200
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10163 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  -
;      LOCATION:  1..10163
;      OTHER INFORMATION:  /note= "HIV-2KR provirus"
US-08-659-251-1

  Query Match             51.0%;  Score 1346.4;  DB 2;  Length 10163;
  Best Local Similarity   72.2%;  Pred. No. 0;
  Matches 1911;  Conservative    0;  Mismatches  651;  Indels   84;  Gaps    9;

Qy       2 CTAGAAGCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTGT 61
Db    6502 CTAGAGGCATGCAATAACCCATGTTATTGTAAGAAATGTTGTTACCATTGCCAGCTATGT 6561

Qy      62 TTTCTTAAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCCG 121
Db    6562 TTTTTAAAAAAGGGACTCGGGATATGTTATGAACGGAAGGGCAGACGAAGAAGGACTC-- 6619

Qy     122 AAAAAGGCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGGA 181
Db    6620 -------CAAGGGCTCATTCGTCTTCTGCATCAGACAAGTGAGTATAATGGATAGTAGAA 6672

Qy     182 ATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATATG 241
Db    6673 ATCAGCTAATTGTTGCCATTTTACTAACTAGTGCTTGCTTAATATATTGCGCCCAATATG 6732

Qy     242 TCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCAA 301
Db    6733 TGACTGTTTTCTATGGCATACCCGCGTGGAAGAATGCATCCATTCCCCTCTTTTGTGCAA 6792

Qy     302 CCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCAG 361
Db    6793 CCAGAAATAGAGATACTTGGGGAACCATACAGTGCTTGCCAGACAATGATGATTATCAGG 6852

Qy     362 AAGTGGCCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAGG 421
Db    6853 AAATACCTTTAAATGTGACAGAGGCTTTTGACGCATGGAACAATACAGTAACAGAACAAG 6912

Qy     422 CAATAGAGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCCC 481
Db    6913 CAGTAGAAGATGTCTGGAATCTATTTGAGACATCAGTAAAACCATGTGTCAAATTAACAC 6972

Qy     482 CATTATGCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAAT 541

Db    6973 CCTTATGTGTGCAAATGGAATGTAACAGCACAAGTACAG---------------------- 7011

Qy     542 CAATAACAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTCA 601
Db    7012 -----------AGAGCAGTAACAGCACAAGTGAGGGGAGCA-CAGTCCCAGAGATATTAA 7059

Qy     602 ATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCAAATGA 661
Db    7060 ACGAAACTACTTCATGCATAACCAACAACAGCTGCTCAGATTTAGGGAGTGAAGAGGTAG 7119

Qy     662 TAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAATGAAA 721
Db    7120 TCGATTGTCGGTTCAATATGACAGGACTACAACTAGATAAGCCACAGCAATATAGTGAAA 7179

Qy     722 CTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGAT 781
Db    7180 CATGGTACTCAAAGGATGTAGTTTGTGA------CACAACTAATGGGACCAGCCGCAAAT 7233

Qy     782 GTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGGG 841
Db    7234 GTTACATGAACCATTGCAACACATCAGTCATCACAGAGTCATGTGATAAGCACTATTGGG 7293

Qy     842 ATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGACA 901
Db    7294 ATGCTATGAGGTTTAGATACTGTGCACCACCGGGTTTATGCTTGCTAAGATGCAATGATA 7353

Qy     902 CAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGGA 961
Db    7354 CCAATTATTCAGGCTTTGAGCCCAAGTGTCCTAAAGTAGTAGCTGCTACATGCACAAGAA 7413

Qy     962 TGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGAA 1021
Db    7414 TGATGGAAACGCAAACTTCTACTTGGTTTGGCTTTAATGGCACTAGGGCAGAAAATAGAA 7473

Qy    1022 CTTATATTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTATA 1081
Db    7474 CATATATCTATTGGCATGGTAGAGATAATAGGACTATTATCAGCTTAAATACACATTATA 7533

Qy    1082 ATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATGT 1141
Db    7534 ATCTCACAATGCATTGTAAGAGGCCAGGAAATAAGTCAGTTTTGCCAATAACACTTAGGT 7593

Qy    1142 CTGGATTGGTTTTCCACTCACAACC---AATCAATGATAGGCCAAAGCAGGCATGGTGTT 1198
Db    7594 CAGGGAGAGTGTTTCACTCCCGACCGATCATCAATGAAAGACCCAAGCAGGCATGGTGCT 7653

Qy    1199 GGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATC 1258
Db    7654 GGTTCGGAGGTGATTGGAAGAAAGCCATGCAGGAGGTGAAACAAACCCTTGTGAAACATC 7713

Qy    1259 CCAGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGAGGAGG-- 1316
Db    7714 CCAGGTATAGAGGAACCAACGACACACAGAAAATTAACTTTACACAACCAGGAAAAGGTT 7773

Qy    1317 -AGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAA 1375
Db    7774 CAGATGCAGAAGTGGTATACATGTGGACTAACTGCAGAGGAGAATTTCTATACTGCAACA 7833

Qy    1376 TGAATTGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGGAAC 1435
Db    7834 TGACTCGGTTCCTCAATTGGATAGAAAACAG----------------------GGCAC 7869

Qy    1436 AGCATAAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAG 1495
Db    7870 ACCCACAGCGCAATTATGCACCGTGCCATATAAGGCAAATAATTAATACCTGGCATAGAG 7929

Qy    1496 TAGGCAAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGA 1555
Db    7930 TAGGCCAAAATATATATTTGCCTCCTAGGGAAGGGGAATTGGTCTGCAACTCAACAGTAA 7989

Qy    1556 CCAGTCTCATAGCAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTG 1615
```

```
Db   7990 CCAGCATAATTGCTAACATTGACATGTTTGAT---AACCAGACAAGCATTACCTTTAGTG 8046

Qy   1616 CAGAGGTGGCAGAACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTC 1675
          ||||||||||||||| || ||||||||||||||||| |||||||||| || || |
Db   8047 CAGAGGTGGCAGAACTATACCGATTGGAATTGGGAGATTACAAATTAGTAGAAATAACAC 8106

Qy   1676 CAATTGGCTTGGCCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATA 1735
          |||||||||| || || ||| | || ||||| | ||| | | || ||||
Db   8107 CAATTGGCTTCGCACCTACATCAGAAAAAAGGTATTCCTCTGCT---CCACAGAGGAATA 8163

Qy   1736 AAAGAGGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGG 1795
          ||||||| || |||||||||||  ||||||||||||||||||| ||||||||||||||||
Db   8164 AAAGAGGTGTGTTTGTGCTAGGAGTCTTGGGTTTTCTCGCAACAGCAGGTTCTGCAATGG 8223

Qy   1796 GCGCGGCGTCGTTGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGC 1855
          |||||||||| |||||||| | |||||  || || ,||  | ||||||||||||||||
Db   8224 GCGCGGCGTCCTTGACGCTGTCGGCTCATCCCGGACTTTACTGGGCTGGGATAGTGCAGC 8283

Qy   1856 AACAGCAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCT 1915
          |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   8284 AACAGCAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAAATGTTGCGACTGACCGTCT 8343

Qy   1916 GGGGAACAAAGAACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGG 1975
          |||||||||| || |||||||| || |||||||| |||||||| ||| ||| ||||||||
Db   8344 GGGGAACAAAAAATCTCCAGACAAGAGTCACTGCTATCGAGAAATACCTAAGGGACCAGG 8403

Qy   1976 CGCAGCTGAATGCTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGC 2035
          ||| ||| ||| | |||||||||| |||||||||||||| |||| ||||||| ||||
Db   8404 CGCGGCTAAATTCATGGGGATGTGCATTTAGACAAGTCTGCTACACCACTGTACTATGGG 8463

Qy   2036 CAAATGCAAGTCTAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGG 2095
          ||||   ||  ||  |||  | ||||||||  ||| ||||| || ||||| | | |
Db   8464 AAAATAACAGCATAGTACCTGATTGGAACAACATGACGTGGCAGGAATGGGAACAACAAA 8523

Qy   2096 TTGACTTCTTGGAAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGA 2155
              | | || | |||||| |         ||||| ||||||||| ||||||||||
Db   8524 CCCGCGACCTAGAGGCAAATATCAGTAGATCGTTAGAGCAGGCACAAATCCAACAAGAGA 8583

Qy   2156 AGAACATGTATGAATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACC 2215
          | || ||||||||  ||||||| || |||||||||||||| ||||||||  ||||||||
Db   8584 AAAATATGTATGAGCTACAAAAATTAAATAGCTGGGATGTTTTTGGCAACTGGTTTGATT 8643

Qy   2216 TTGCTTCTTGGATAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGT 2275
          |  | || ||||| |||||||| || |||||||||||| || |  ||||| |||||   |
Db   8644 TAACCTCCTGGATTAAGTATATTCAGTATGGAGTTTATGTAATAATAGGAATAATAGCTT 8703

Qy   2276 TAAGAATAGTGATCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAG 2335
          |||||||||| || ||| |||||||| | |||  ||  | ||  |||| |||||||| |
Db   8704 TAAGAATAGTAATATATGTAGTACAATTACTAAGTAGACTTAGAAAGGGCTATAGGCCTG 8763

Qy   2336 TGTTCTCTTCCCCACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGC 2395
          | |||||||||| ||| |||| |||| |||| |||||||||  |||||  || || ||
Db   8764 TTTTCTCTTCCCCCCCCGGTTATATCCAACAGATCCATATCCACAAGGACTGGGAACAGC 8823

Qy   2396 CAACCAGAGAAGGCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTT 2455
          ||  |||||||  | |||  |||||  ||| | ||  |||  ||||||| |||||||
Db   8824 CAGACAGAGAAGAAACAGACGAAGACGCCGGAAACAGCATTGGAGACAGCTCGTGGCCTT 8883

Qy   2456 GGCAGATAGAATATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTAT 2515
          |||  |||| |||||| |||||||||||||||| ||||| ||||||||||  ||||||
Db   8884 GGCCAATAGCATATATACATTTCCTGATCCGCCAGCTGATTCGCCTCTTGACCGGGCTAT 8943

Qy   2516 TCAGCAACTGCAGAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGA 2575
           ||||  |||||    ||| ||||| ||    | || || |||||||| |||  ||||||
Db   8944 ACAGCGTCTGCAAGGACTTACTATCCAGGAGCTTCCCGACCCTCCAACTAATCTTCCAGA 9003

Qy   2576 GGCTCTCTGCGACCCTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTACCTAC 2635
          | ||         | |||   |  ||  | ||    || ||   |  |  |||||||  |
Db   9004 GTCTTCAGAGAGCACTAACAACAATCAGGGACTGGCTGAGACTTACAATAGCCTACCTGC 9063

Qy   2636 AATATG 2641
          ||||||
Db   9064 AATATG 9069
```

```
RESULT 7
US-09-256-490-1
; Sequence 1, Application US/09256490
; Patent No. 6235881
;  GENERAL INFORMATION:
;    APPLICANT:  Kraus, Guenter
;    APPLICANT:  Wong-Staal, Flossie
;    APPLICANT:  Talbott, Randy
;    APPLICANT:  Poeschla, Eric
;    TITLE OF INVENTION:  Isolation of No. 6235881el HIV-2 Proviruses
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew LLP
;      STREET:  Two Embarcadero Center, Eighth Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/256,490
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/659,251
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Garrett-Wackowski, Eugenia
;      REGISTRATION NUMBER:  37,330
;      REFERENCE/DOCKET NUMBER:  02307E-056410US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 576-0200
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10163 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  -
;      LOCATION:  1..10163
;      OTHER INFORMATION:  /note= "HIV-2KR provirus"
US-09-256-490-1

  Query Match             51.0%;  Score 1346.4;  DB 4;  Length 10163;
  Best Local Similarity   72.2%;  Pred. No. 0;
  Matches 1911;  Conservative    0;  Mismatches  651;  Indels   84;  Gaps    9;

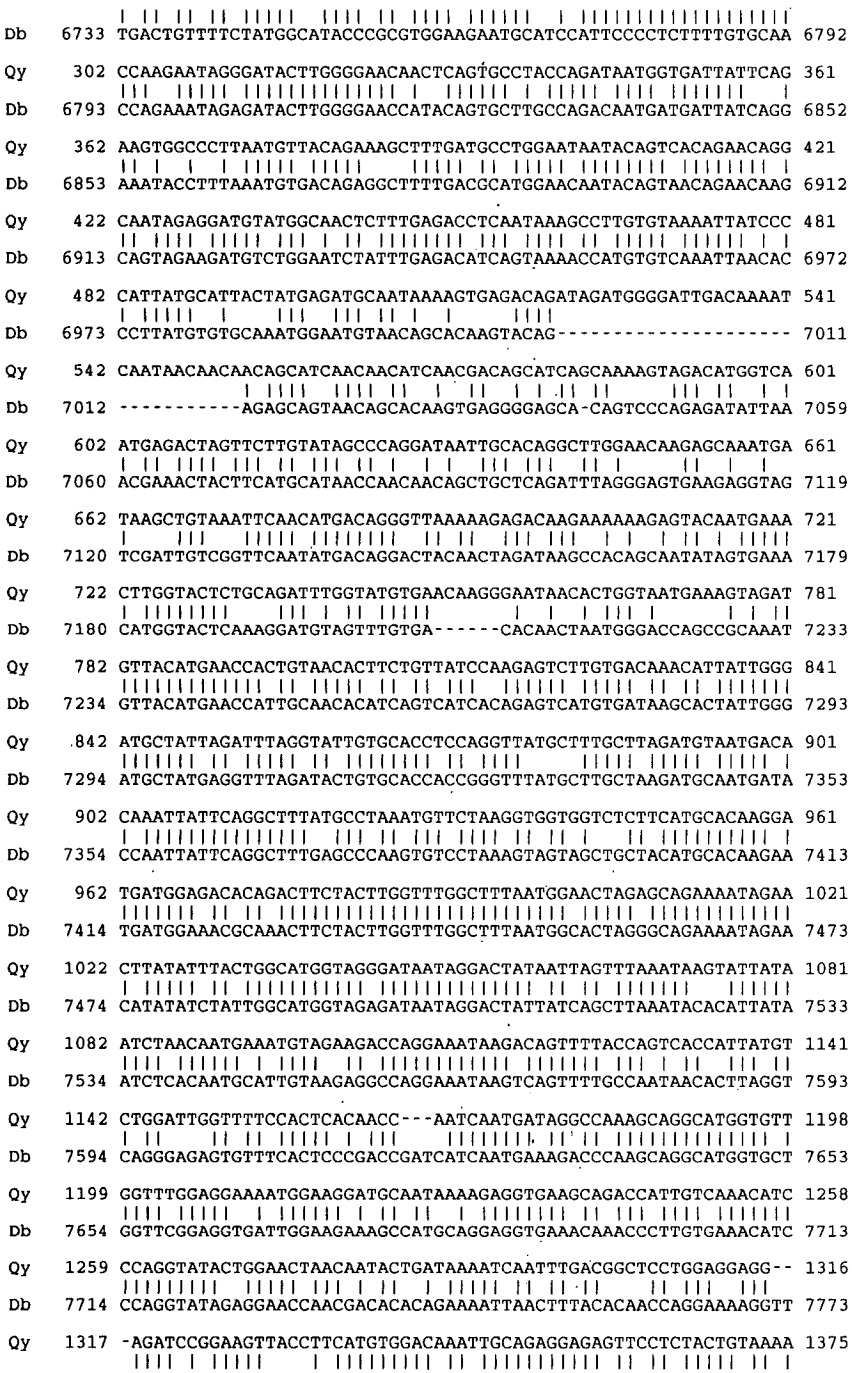
Qy       2 CTAGAAGCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTGT 61
           ||||| |||||| ||||| |||| |||||||| || ||||| ||||||||||| | |||
Db    6502 CTAGAGGCATGCAATAACCCATGTTATTGTAAGAAATGTTGTTACCATTGCCAGCTATGT 6561

Qy      62 TTTCTTAAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCCG 121
           ||| | ||||| ||  | |||||||||||||| |     | |   ||||||| ||||
Db    6562 TTTTTAAAAAAGGGACTCGGGATATGTTATGAACGGAAGGGCAGACGAAGAAGGACTC-- 6619

Qy     122 AAAAAGGCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGGA 181
                 | | |||| || | |||||||||||| |||||| |||||      | | | |
Db    6620 -------CAAGGGCTCATTCGTCTTCTGCATCAGACAAGTGAGTATAATGGATAGTAGAA 6672

Qy     182 ATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATATG 241
           |||||||  || | ||||| || ||    |||| |     || |||||  | | |||||
Db    6673 ATCAGCTAATTGTTGCCATTTTACTAACTAGTGCTTGCTTAATATATTGCGCCCAATATG 6732

Qy     242 TCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCAA 301
```


Db 6733 TGACTGTTTTCTATGGCATACCCGCGTGGAAGAATGCATCCATTCCCCTCTTTTGTGCAA 6792
Qy 302 CCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCAG 361
Db 6793 CCAGAAATAGAGATACTTGGGGAACCATACAGTGCTTGCCAGACAATGATGATTATCAGG 6852
Qy 362 AAGTGGCCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAGG 421
Db 6853 AAATACCTTTAAATGTGACAGAGGCTTTTGACGCATGGAACAATACAGTAACAGAACAAG 6912
Qy 422 CAATAGAGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCCC 481
Db 6913 CAGTAGAAGATGTCTGGAATCTATTTGAGACATCAGTAAAACCATGTGTCAAATTAACAC 6972
Qy 482 CATTATGCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAAT 541
Db 6973 CCTTATGTGTGCAAATGGAATGTAACAGCACAAGTACAG---------- 7011
Qy 542 CAATAACAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTCA 601
Db 7012 ----------AGAGCAGTAACAGCACAAGTGAGGGGAGCA-CAGTCCCAGAGATATTAA 7059
Qy 602 ATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCAAATGA 661
Db 7060 ACGAAACTACTTCATGCATAACCAACAACAGCTGCTCAGATTTAGGGAGTGAAGAGGTAG 7119
Qy 662 TAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAATGAAA 721
Db 7120 TCGATTGTCGGTTCAATATGACAGGACTACAACTAGATAAGCCACAGCAATATAGTGAAA 7179
Qy 722 CTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGAT 781
Db 7180 CATGGTACTCAAAGGATGTAGTTTGTGA------CACAACTAATGGGACCAGCCGCAAAT 7233
Qy 782 GTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGGG 841
Db 7234 GTTACATGAACCATTGCAACACATCAGTCATCACAGAGTCATGTGATAAGCACTATTGGG 7293
Qy 842 ATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGACA 901
Db 7294 ATGCTATGAGGTTTAGATACTGTGCACCACCGGGTTTATGCTTGCTAAGATGCAATGATA 7353
Qy 902 CAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGGA 961
Db 7354 CCAATTATTCAGGCTTTGAGCCCAAGTGTCCTAAAGTAGTAGCTGCTACATGCACAAGAA 7413
Qy 962 TGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGAA 1021
Db 7414 TGATGGAAACGCAAACTTCTACTTGGTTTGGCTTTAATGGCACTAGGGCAGAAAATAGAA 7473
Qy 1022 CTTATATTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTATA 1081
Db 7474 CATATATCTATTGGCATGGTAGAGATAATAGGACTATTATCAGCTTAAATACACATTATA 7533
Qy 1082 ATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATGT 1141
Db 7534 ATCTCACAATGCATTGTAAGAGGCCAGGAAATAAGTCAGTTTTGCCAATAACACTTAGGT 7593
Qy 1142 CTGGATTGGTTTTCCACTCACAACC---AATCAATGATAGGCCAAAGCAGGCATGGTGTT 1198
Db 7594 CAGGGAGAGTGTTTCACTCCCGACCGATCATCAATGAAAGACCCAAGCAGGCATGGTGCT 7653
Qy 1199 GGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATC 1258
Db 7654 GGTTCGGAGGTGATTGGAAGAAAGCCATGCAGGAGGTGAAACAAACCCTTGTGAAACATC 7713
Qy 1259 CCAGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGAGGAGG-- 1316
Db 7714 CCAGGTATAGAGGAACCAACGACACACAGAAAATTAACTTTACACAACCAGGAAAAGGTT 7773
Qy 1317 -AGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAA 1375
Db 7774 CAGATGCAGAAGTGGTATACATGTGGACTAACTGCAGAGGAGAATTTCTATACTGCAACA 7833
Qy 1376 TGAATTGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGGAAC 1435
Db 7834 TGACTCGGTTCCTCAATTGGATAGAAAACAG----------GGCAC 7869
Qy 1436 AGCATAAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAG 1495
Db 7870 ACCCACAGCGCAATTATGCACCGTGCCATATAAGGCAAATAATTAATACCTGGCATAGAG 7929
Qy 1496 TAGGCAAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGA 1555
Db 7930 TAGGCCAAAATATATATTTGCCTCCTAGGGAAGGGGAATTGGTCTGCAACTCAACAGTAA 7989
Qy 1556 CCAGTCTCATAGCAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTG 1615
Db 7990 CCAGCATAATTGCTAACATTGACATGTTTGAT---AACCAGACAAGCATTACCTTTAGTG 8046
Qy 1616 CAGAGGTGGCAGAACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTC 1675
Db 8047 CAGAGGTGGCAGAACTATACCGATTGGAATTGGGAGATTACAAATTAGTAGAAATAACAC 8106
Qy 1676 CAATTGGCTTGGCCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATA 1735
Db 8107 CAATTGGCTTCGCACCTACATCAGAAAAAAGGTATTCCTCTGCT---CCACAGAGGAATA 8163
Qy 1736 AAAGAGGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGG 1795
Db 8164 AAAGAGGTGTGTTTGTGCTAGGAGTCTTGGGTTTTCTCGCAACAGCAGGTTCTGCAATGG 8223
Qy 1796 GCGCGGCGTCGTTGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGC 1855
Db 8224 GCGCGGCGTCCTTGACGCTGTCGGCTCATCCCGGACTTTACTGGGCTGGGATAGTGCAGC 8283
Qy 1856 AACAGCAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCT 1915
Db 8284 AACAGCAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAAATGTTGCGACTGACCGTCT 8343
Qy 1916 GGGGAACAAAGAACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGG 1975
Db 8344 GGGGAACAAAAAATCTCCAGACAAGAGTCACTGCTATCGAGAAATACCTAAGGGACCAGG 8403
Qy 1976 CGCAGCTGAATGCTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGC 2035
Db 8404 CGCGGCTAAATTCATGGGGATGTGCATTTAGACAAGTCTGCTACACCACTGTACTATGGG 8463
Qy 2036 CAAATGCAAGTCTAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGG 2095
Db 8464 AAAATAACAGCATAGTACCTGATTGGAACAACATGACGTGGCAGGAATGGGAACAACAAA 8523
Qy 2096 TTGACTTCTTGGAAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGA 2155
Db 8524 CCCGCGACCTAGAGGCAAATATCAGTAGATCGTTAGAGCAGGCACAAATCCAACAAGAGA 8583
Qy 2156 AGAACATGTATGAATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACC 2215
Db 8584 AAAATATGTATGAGCTACAAAAATTAAATAGCTGGGATGTTTTTGGCAACTGGTTTGATT 8643
Qy 2216 TTGCTTCTTGGATAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGT 2275
Db 8644 TAACCTCCTGGATTAAGTATATTCAGTATGGAGTTTATGTAATAATAGGAATAATAGCTT 8703
Qy 2276 TAAGAATAGTGATCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAG 2335
Db 8704 TAAGAATAGTAATATATGTAGTACAATTACTAAGTAGACTTAGAAAGGGCTATAGGCCTG 8763
Qy 2336 TGTTCTCTTCCCCACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGC 2395
Db 8764 TTTTCTCTTCCCCCCCCGGTTATATCCAACAGATCCATATCCACAAGGACTGGGAACAGC 8823
Qy 2396 CAACCAGAGAAGGCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTT 2455
Db 8824 CAGACAGAGAAGAAACAGACGAAGACGCCGGAAACAGCATTGGAGACAGCTCGTGGCCTT 8883

```
Qy   2456 GGCAGATAGAATATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTAT 2515
          |||  |||| |||||| ||||||||||||||||| |||| |||||||||||  ||||||
Db   8884 GGCCAATAGCATATATACATTTCCTGATCCGCCAGCTGATTCGCCTCTTGACCGGGCTAT 8943

Qy   2516 TCAGCAACTGCAGAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGA 2575
           ||||  |||||    ||| ||||| ||   || || || |||||||| |||  ||||||
Db   8944 ACAGCGTCTGCAAGGACTTACTATCCAGGAGCTTCCCGACCCTCCAACTAATCTTCCAGA 9003

Qy   2576 GGCTCTCTGCGACCCTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTACCTAC 2635
          | ||        | |||   |  ||  | ||  || ||   |  | |  |||||||  |
Db   9004 GTCTTCAGAGAGCACTAACAACAATCAGGGACTGGCTGAGACTTACAATAGCCTACCTGC 9063

Qy   2636 AATATG 2641
          ||||||
Db   9064 AATATG 9069


RESULT 8
PCT-US96-11445-1
; Sequence 1, Application PC/TUS9611445
;  GENERAL INFORMATION:
;    APPLICANT:  The Regents of the University of California
;    TITLE OF INVENTION:  Isolation of Novel HIV-2 Proviruses
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Robbins, Berliner & Carson
;      STREET:  201 N. Figueroa Street, 5th Floor
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  90012-2628
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/11445
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berliner, Robert
;      REGISTRATION NUMBER:  20,121
;      REFERENCE/DOCKET NUMBER:  5555-399C1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 977-1001
;      TELEFAX:  (213) 977-1003
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10163 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  -
;      LOCATION:  1..10163
;      OTHER INFORMATION:  /note= "HIV-2KR provirus"
PCT-US96-11445-1

  Query Match             51.0%;  Score 1346.4;  DB 5;  Length 10163;
  Best Local Similarity   72.2%;  Pred. No. 0;
  Matches 1911;  Conservative    0;  Mismatches  651;  Indels   84;  Gaps    9;

Qy       2 CTAGAAGCATGCTATAACACATGCTATTGTAAAAAGTGTTGCTACCATTGCCAGTTTTGT 61
           ||||| |||||| ||||| |||| |||||||| || ||||| ||||||||||| | |||
Db    6502 CTAGAGGCATGCAATAACCCATGTTATTGTAAGAAATGTTGTTACCATTGCCAGCTATGT 6561

Qy      62 TTTCTTAAAAAAGGCTTGGGGATATGTTATGAGCAATCACGAAAGAGAAGAAGAACTCCG 121
           ||| | ||||| ||  | ||||||||||||| |      | |   ||||||| ||||
Db    6562 TTTTTAAAAAAGGGACTCGGGATATGTTATGAACGGAAGGGCAGACGAAGAAGGACTC-- 6619

Qy     122 AAAAAGGCTAAGGCTAATACATCTTCTGCATCAAACAAGTAAGTATGGGATGTCTTGGGA 181
                 | | |||| || | |||||||||||| |||||| |||||      |  | | |
Db    6620 -------CAAGGGCTCATTCGTCTTCTGCATCAGACAAGTGAGTATAATGGATAGTAGAA 6672

Qy     182 ATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCTATATG 241
           |||||||  || | ||||| || ||    ||||  |     || |||||  | | |||||
Db    6673 ATCAGCTAATTGTTGCCATTTTACTAACTAGTGCTTGCTTAATATATTGCGCCCAATATG 6732

Qy     242 TCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTGTGCAA 301
           | || || || |||||  |||| || |||| ||||||  | ||||||||||||||||||
Db    6733 TGACTGTTTTCTATGGCATACCCGCGTGGAAGAATGCATCCATTCCCCTCTTTTGTGCAA 6792

Qy     302 CCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCAG 361
           |||  ||||| |||||||||||||| |  |||||| | ||||| |||| |||||||   |
Db    6793 CCAGAAATAGAGATACTTGGGGAACCATACAGTGCTTGCCAGACAATGATGATTATCAGG 6852

Qy     362 AAGTGGCCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAGG 421
           || |  |  | ||||| |||||    ||||| || ||||| |||||||| |||||||| |
Db    6853 AAATACCTTTAAATGTGACAGAGGCTTTTGACGCATGGAACAATACAGTAACAGAACAAG 6912

Qy     422 CAATAGAGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCCC 481
           || |||| ||||| ||| | || |||||||| ||| |||| || ||||| |||||| | |
Db    6913 CAGTAGAAGATGTCTGGAATCTATTTGAGACATCAGTAAAACCATGTGTCAAATTAACAC 6972

Qy     482 CATTATGCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAAT 541
           | |||||  |    |||  ||| || |  |     ||||
Db    6973 CCTTATGTGTGCAAATGGAATGTAACAGCACAAGTACAG---------------------- 7011

Qy     542 CAATAACAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTCA 601
                   | ||||  |||| ||  |  ||  | | || ||     ||| ||  | |
Db    7012 ----------AGAGCAGTAACAGCACAAGTGAGGGGAGCA-CAGTCCCAGAGATATTAA 7059

Qy     602 ATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCAAATGA 661
           | || |||| ||| || ||| || |  | |  ||| |||  || |     ||  |  |
Db    7060 ACGAAACTACTTCATGCATAACCAACAACAGCTGCTCAGATTTAGGGAGTGAAGAGGTAG 7119

Qy     662 TAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAATGAAA 721
           |    |||   ||||| |||||||| || ||  ||| ||| | |  | || | |||||
Db    7120 TCGATTGTCGGTTCAATATGACAGGACTACAACTAGATAAGCCACAGCAATATAGTGAAA 7179

Qy     722 CTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGAT 781
           | ||||||||    ||| | || |||||       |  |  | ||| |      | | ||
Db    7180 CATGGTACTCAAAGGATGTAGTTTGTGA------CACAACTAATGGGACCAGCCGCAAAT 7233

Qy     782 GTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGGG 841
           |||||||||||| || ||||| || || ||| |||||| |||||  || || |||||||
Db    7234 GTTACATGAACCATTGCAACACATCAGTCATCACAGAGTCATGTGATAAGCACTATTGGG 7293

Qy     842 ATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGACA 901
           ||||||| || ||||| || |||||||| || ||||     ||||| ||||| ||||| |
Db    7294 ATGCTATGAGGTTTAGATACTGTGCACCACCGGGTTTATGCTTGCTAAGATGCAATGATA 7353

Qy     902 CAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGGA 961
           | ||||||||||||||  ||| || ||| |||| || || |   || ||||||||| |
Db    7354 CCAATTATTCAGGCTTTGAGCCCAAGTGTCCTAAAGTAGTAGCTGCTACATGCACAAGAA 7413

Qy     962 TGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAATAGAA 1021
           ||||||| || || |||||||||||||||||||||||||| ||||| ||||||||||||
Db    7414 TGATGGAAACGCAAACTTCTACTTGGTTTGGCTTTAATGGCACTAGGGCAGAAAATAGAA 7473

Qy    1022 CTTATATTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTATA 1081
           | ||||| || |||||||||||| |||||||||||||| || || |||||||   ||||||
Db    7474 CATATATCTATTGGCATGGTAGAGATAATAGGACTATTATCAGCTTAAATACACATTATA 7533

Qy    1082 ATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATGT 1141
           |||| |||||| | |||  || |||||||||||| ||||||| ||| | ||  ||| ||
Db    7534 ATCTCACAATGCATTGTAAGAGGCCAGGAAATAAGTCAGTTTTGCCAATAACACTTAGGT 7593

Qy    1142 CTGGATTGGTTTTCCACTCACAACC---AATCAATGATAGGCCAAAGCAGGCATGGTGTT 1198
           | ||    || || ||||| | |||    |||||||| || || ||||||||||||| |
Db    7594 CAGGGAGAGTGTTTCACTCCCGACCGATCATCAATGAAAGACCCAAGCAGGCATGGTGCT 7653

Qy    1199 GGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATC 1258
```

```
Db   7654 GGTTCGGAGGTGATTGGAAGAAAGCCATGCAGGAGGTGAAACAAACCCTTGTGAAACATC 7713

Qy   1259 CCAGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGAGGAGG-- 1316

Db   7714 CCAGGTATAGAGGAACCAACGACACACAGAAAATTAACTTTACACAACCAGGAAAAGGTT 7773

Qy   1317 -AGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAA 1375

Db   7774 CAGATGCAGAAGTGGTATACATGTGGACTAACTGCAGAGGAGAATTTCTATACTGCAACA 7833

Qy   1376 TGAATTGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGGAAC 1435

Db   7834 TGACTCGGTTCCTCAATTGGATAGAAAACAG----------GGCAC 7869

Qy   1436 AGCATAAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAG 1495

Db   7870 ACCCACAGCGCAATTATGCACCGTGCCATATAAGGCAAATAATTAATACCTGGCATAGAG 7929

Qy   1496 TAGGCAAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGA 1555

Db   7930 TAGGCCAAAATATATATTTGCCTCCTAGGGAAGGGGAATTGGTCTGCAACTCAACAGTAA 7989

Qy   1556 CCAGTCTCATAGCAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATGAGTG 1615

Db   7990 CCAGCATAATTGCTAACATTGACATGTTTGAT---AACCAGACAAGCATTACCTTTAGTG 8046

Qy   1616 CAGAGGTGGCAGAACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGATCACTC 1675

Db   8047 CAGAGGTGGCAGAACTATACCGATTGGAATTGGGAGATTACAAATTAGTAGAAATAACAC 8106

Qy   1676 CAATTGGCTTGGCCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATA 1735

Db   8107 CAATTGGCTTCGCACCTACATCAGAAAAAAGGTATTCCTCTGCT---CCACAGAGGAATA 8163

Qy   1736 AAAGAGGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGG 1795

Db   8164 AAAGAGGTGTGTTTGTGCTAGGAGTCTTGGGTTTTCTCGCAACAGCAGGTTCTGCAATGG 8223

Qy   1796 GCGCGGCGTCGTTGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGTGCAGC 1855

Db   8224 GCGCGGCGTCCTTGACGCTGTCGGCTCATCCCGGACTTTACTGGGCTGGGATAGTGCAGC 8283

Qy   1856 AACAGCAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCT 1915

Db   8284 AACAGCAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAAATGTTGCGACTGACCGTCT 8343

Qy   1916 GGGGAACAAAGAACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGACCAGG 1975

Db   8344 GGGGAACAAAAAATCTCCAGACAAGAGTCACTGCTATCGAGAAATACCTAAGGGACCAGG 8403

Qy   1976 CGCAGCTGAATGCTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCATGGC 2035

Db   8404 CGCGGCTAAATTCATGGGGATGTGCATTTAGACAAGTCTGCTACACCACTGTACTATGGG 8463

Qy   2036 CAAATGCAAGTCTAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGG 2095

Db   8464 AAAATAACAGCATAGTACCTGATTGGAACAACATGACGTGGCAGGAATGGGAACAACAAA 8523

Qy   2096 TTGACTTCTTGGAAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGA 2155

Db   8524 CCCGCGACCTAGAGGCAAATATCAGTAGATCGTTAGAGCAGGCACAAATCCAACAAGAGA 8583

Qy   2156 AGAACATGTATGAATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACC 2215

Db   8584 AAAATATGTATGAGCTACAAAAATTAAATAGCTGGGATGTTTTTGGCAACTGGTTTGATT 8643

Qy   2216 TTGCTTCTTGGATAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGT 2275

Db   8644 TAACCTCCTGGATTAAGTATATTCAGTATGGAGTTTATGTAATAATAGGAATAATAGCTT 8703

Qy   2276 TAAGAATAGTGATCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAG 2335

Db   8704 TAAGAATAGTAATATATGTAGTACAATTACTAAGTAGACTTAGAAAGGGCTATAGGCCTG 8763

Qy   2336 TGTTCTCTTCCCCACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGC 2395

Db   8764 TTTTCTCTTCCCCCCCCGGTTATATCCAACAGATCCATATCCACAAGGACTGGGAACAGC 8823

Qy   2396 CAACCAGAGAAGGCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTT 2455

Db   8824 CAGACAGAGAAGAAACAGACGAAGACGCCGGAAACAGCATTGGAGACAGCTCGTGGCCTT 8883

Qy   2456 GGCAGATAGAATATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTAT 2515

Db   8884 GGCCAATAGCATATATACATTTCCTGATCCGCCAGCTGATTCGCCTCTTGACCGGGCTAT 8943

Qy   2516 TCAGCAACTGCAGAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGA 2575

Db   8944 ACAGCGTCTGCAAGGACTTACTATCCAGGAGCTTCCCGACCCTCCAACTAATCTTCCAGA 9003

Qy   2576 GGCTCTCTGCGACCCTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTACCTAC 2635

Db   9004 GTCTTCAGAGAGCACTAACAACAATCAGGGACTGGCTGAGACTTACAATAGCCTACCTGC 9063

Qy   2636 AATATG 2641

Db   9064 AATATG 9069
```

```
RESULT 9
US-08-659-251-13
; Sequence 13, Application US/08659251
; Patent No. 5883081
;  GENERAL INFORMATION:
;    APPLICANT:  Kraus, Guenter
;    APPLICANT:  Wong-Staal, Flossie
;    APPLICANT:  Talbott, Randy
;    APPLICANT:  Poeschla, Eric
;    TITLE OF INVENTION:  Isolation of No. 5883081el HIV-2 Proviruses
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew LLP
;      STREET:  Two Embarcadero Center, Eighth Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/659,251
;      FILING DATE:  No. 5883081 yet assigned
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/001,441
;      FILING DATE:  26-JUL-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Garrett-Wackowski, Eugenia
;      REGISTRATION NUMBER:  37,330
;      REFERENCE/DOCKET NUMBER:  02307E-056410US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 576-0200
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2574 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    FEATURE:
;      NAME/KEY:  -
```

```
;       LOCATION:  1..2574
;       OTHER INFORMATION:  /label= env
;       OTHER INFORMATION:  /note= "HIV-2KR subsequence encoding
;       OTHER INFORMATION:  env gene"
US-08-659-251-13

  Query Match             48.0%;  Score 1266.6;  DB 2;  Length 2574;
  Best Local Similarity   72.3%;  Pred. No. 0;
  Matches 1787;  Conservative    0;  Mismatches  609;  Indels   75;  Gaps    8;

Qy    177 TGGGAATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCT 236
          | | ||||||||  || | ||||| || ||    ||||  |     || |||||  | |
Db      9 TAGAAATCAGCTAATTGTTGCCATTTTACTAACTAGTGCTTGCTTAATATATTGCGCCCA 68

Qy    237 ATATGTCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTG 296
          ||||||  || || || |||||  |||| || |||| ||||||  | ||||||||||||||
Db     69 ATATGTGACTGTTTTCTATGGCATACCCGCGTGGAAGAATGCATCCATTCCCCTCTTTTG 128

Qy    297 TGCAACCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTA 356
          ||||||||  ||||| |||||||||||||| |  |||||| | ||||| |||| ||||||
Db    129 TGCAACCAGAAATAGAGATACTTGGGGAACCATACAGTGCTTGCCAGACAATGATGATTA 188

Qy    357 TTCAGAAGTGGCCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGA 416
          |   ||| |  |  | ||||| |||||    ||||| || ||||| |||||||| |||||
Db    189 TCAGGAAATACCTTTAAATGTGACAGAGGCTTTTGACGCATGGAACAATACAGTAACAGA 248

Qy    417 ACAGGCAATAGAGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATT 476
          ||| ||| |||| ||||| ||| | || |||||||| ||| |||| || ||||| |||||
Db    249 ACAAGCAGTAGAAGATGTCTGGAATCTATTTGAGACATCAGTAAAACCATGTGTCAAATT 308

Qy    477 ATCCCCATTATGCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGAC 536
          | | || |||||  |    |||  ||| || |  |     ||||
Db    309 AACACCCTTATGTGTGCAAATGGAATGTAACAGCACAAGTACAG---------------- 352

Qy    537 AAAATCAATAACAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACAT 596
                      | ||||  |||| ||  |  ||  | | || ||     ||| ||
Db    353 ----------------AGAGCAGTAACAGCACAAGTGAGGGGAGCA-CAGTCCCAGAGAT 395

Qy    597 GGTCAATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCA 656
            | || || |||| ||| || ||| || |  | |  ||| |||  || |      ||  |
Db    396 ATTAAACGAAACTACTTCATGCATAACCAACAACAGCTGCTCAGATTTAGGGAGTGAAGA 455

Qy    657 AATGATAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAA 716
            |  |    |||   ||||| ||||||||  || ||  ||| |||  | |  | || |
Db    456 GGTAGTCGATTGTCGGTTCAATATGACAGGACTACAACTAGATAAGCCACAGCAATATAG 515

Qy    717 TGAAACTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAG 776
          ||||||  ||||||||     ||| | || |||||       |  |  | ||| |      |
Db    516 TGAAACATGGTACTCAAAGGATGTAGTTTGTGA------CACAACTAATGGGACCAGCCG 569

Qy    777 TAGATGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTA 836
            | |||||||||||||| || ||||| || || |||  ||||||  ||||| || || ||
Db    570 CAAATGTTACATGAACCATTGCAACACATCAGTCATCACAGAGTCATGTGATAAGCACTA 629

Qy    837 TTGGGATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAA 896
          ||||||||||||  || ||||| || |||||||| || ||||     |||||  ||||| ||
Db    630 TTGGGATGCTATGAGGTTTAGATACTGTGCACCACCGGGTTTATGCTTGCTAAGATGCAA 689

Qy    897 TGACACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCAC 956
          ||| || ||||||||||||||||  ||| || ||| |||| || || |   || |||||||
Db    690 TGATACCAATTATTCAGGCTTTGAGCCCAAGTGTCCTAAAGTAGTAGCTGCTACATGCAC 749

Qy    957 AAGGATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAA 1016
          ||| |||||||| || || ||||||||||||||||||||||||| ||||| ||||||||
Db    750 AAGAATGATGGAAACGCAAACTTCTACTTGGTTTGGCTTTAATGGCACTAGGGCAGAAAA 809

Qy   1017 TAGAACTTATATTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTA 1076
          ||||||  ||||| || ||||||||||| ||||||||||||| || || |||||||   |
Db    810 TAGAACATATATCTATTGGCATGGTAGAGATAATAGGACTATTATCAGCTTAAATACACA 869

Qy   1077 TTATAATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCAT 1136
          |||||||| |||||| | ||||  || |||||||||||| |||||| ||| | ||  |
Db    870 TTATAATCTCACAATGCATTGTAAGAGGCCAGGAAATAAGTCAGTTTTGCCAATAACACT 929

Qy   1137 TATGTCTGGATTGGTTTTCCACTCACAACC---AATCAATGATAGGCCAAAGCAGGCATG 1193
          || ||| ||    || || ||||| | |||    |||||||| || || ||||||||||
Db    930 TAGGTCAGGGAGAGTGTTTCACTCCCGACCGATCATCAATGAAAGACCCAAGCAGGCATG 989

Qy   1194 GTGTTGGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAA 1253
          ||| ||||| |||||  | |||||| | || ||  | |||||||| || ||| |||| ||
Db    990 GTGCTGGTTCGGAGGTGATTGGAAGAAAGCCATGCAGGAGGTGAAACAAACCCTTGTGAA 1049

Qy   1254 ACATCCCAGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGAGG 1313
          |||||||||||||  ||||| ||| | ||  | ||||| || || ||    || |||
Db    1050 ACATCCCAGGTATAGAGGAACCAACGACACACAGAAAATTAACTTTACACAACCAGGAAA 1109

Qy   1314 AGG---AGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTG 1370
          |||   |||| | |||||    | |||||||||| || |||||||||| || || |||||
Db    1110 AGGTTCAGATGCAGAAGTGGTATACATGTGGACTAACTGCAGAGGAGAATTTCTATACTG 1169

Qy   1371 TAAAATGAATTGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAA 1430
           || ||| | |||| || |||||| ||||| | ||
Db    1170 CAACATGACTCGGTTCCTCAATTGGATAGAAAACAG---------------------- 1205

Qy   1431 GGAACAGCATAAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCA 1490
          || ||| |   |  | ||||| |  || || ||||| || |||||||| || || |||||
Db    1206 GGCACACCCACAGCGCAATTATGCACCGTGCCATATAAGGCAAATAATTAATACCTGGCA 1265

Qy   1491 TAAAGTAGGCAAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCAC 1550
          || ||||||| ||||| | |||||||||||| || || || ||  |    || ||||| ||
Db    1266 TAGAGTAGGCCAAAATATATATTTGCCTCCTAGGGAAGGGGAATTGGTCTGCAACTCAAC 1325

Qy   1551 AGTGACCAGTCTCATAGCAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCAT 1610
          ||| ||||| | || || ||||| ||   | |||||   ||||| || |  || ||| |
Db    1326 AGTAACCAGCATAATTGCTAACATTGACATGTTTGAT---AACCAGACAAGCATTACCTT 1382

Qy   1611 GAGTGCAGAGGTGGCAGAACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGAT 1670
          |||||||||||||||||||| || ||||||||||||||||||| ||||||||||  ||
Db    1383 TAGTGCAGAGGTGGCAGAACTATACCGATTGGAATTGGGAGATTACAAATTAGTAGAAAT 1442

Qy   1671 CACTCCAATTGGCTTGGCCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAG 1730
          || |||||||||| || || |||   |  || |||||  |  ||| |   |    ||
Db    1443 AACACCAATTGGCTTCGCACCTACATCAGAAAAAAGGTATTCCTCTGCT---CCACAGAG 1499

Qy   1731 AAATAAAAGAGGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGC 1790
          |||||||||| || |||||||||| |||||||||||||||||| ||||||||||
Db    1500 GAATAAAAGAGGTGTGTTTGTGCTAGGAGTCTTGGGTTTTCTCGCAACAGCAGGTTCTGC 1559

Qy   1791 AATGGGCGCGGCGTCGTTGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGT 1850
          |||||||||||||| |||||||| | |||||  || ||  ||  | ||||||||||||
Db    1560 AATGGGCGCGGCGTCCTTGACGCTGTCGGCTCATCCCGGACTTTACTGGGCTGGGATAGT 1619

Qy   1851 GCAGCAACAGCAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGAC 1910
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    1620 GCAGCAACAGCAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAAATGTTGCGACTGAC 1679

Qy   1911 CGTCTGGGGAACAAAGAACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGA 1970
          |||||||||||||| || |||||||| |||||||| |||||||| ||| ||| |||
Db    1680 CGTCTGGGGAACAAAAAATCTCCAGACAAGAGTCACTGCTATCGAGAAATACCTAAGGGA 1739

Qy   1971 CCAGGCGCAGCTGAATGCTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACC 2030
          |||||||| ||| ||| | |||||||||||| |||||||||||||| |||| |||||||
Db    1740 CCAGGCGCGGCTAAATTCATGGGGATGTGCATTTAGACAAGTCTGCTACACCACTGTACT 1799

Qy   2031 ATGGCCAAATGCAAGTCTAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCG 2090
          ||||  ||||   ||  ||  |||  | ||||||||   ||| ||||| || ||||| |
Db    1800 ATGGGAAAATAACAGCATAGTACCTGATTGGAACAACATGACGTGGCAGGAATGGGAACA 1859

Qy   2091 AAAGGTTGACTTCTTGGAAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACA 2150
          | |      | | | || | |||||| |          ||||| |||||||||| |||||
Db    1860 ACAAACCCGCGACCTAGAGGCAAATATCAGTAGATCGTTAGAGCAGGCACAAATCCAACA 1919

Qy   2151 AGAGAAGAACATGTATGAATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTT 2210
          |||||| || |||||||  |||||| || |||||||||||| |||||||| |||||
Db    1920 AGAGAAAAATATGTATGAGCTACAAAAATTAAATAGCTGGGATGTTTTTGGCAACTGGTT 1979
```

```
Qy    2211 TGACCTTGCTTCTTGGATAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAAT 2270
           |||  |  | || ||||| |||||||| || ||||||||||| || |  ||||| ||||
Db    1980 TGATTTAACCTCCTGGATTAAGTATATTCAGTATGGAGTTTATGTAATAATAGGAATAAT 2039

Qy    2271 ACTGTTAAGAATAGTGATCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAG 2330
           |  ||||||||||| || ||| |||||||| | |||  ||   | ||  |||| |||||
Db    2040 AGCTTTAAGAATAGTAATATATGTAGTACAATTACTAAGTAGACTTAGAAAGGGCTATAG 2099

Qy    2331 GCCAGTGTTCTCTTCCCCACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGC 2390
           ||| || |||||||||||| |||  |||| |||| |||| |||||||||  |||||  ||
Db    2100 GCCTGTTTTCTCTTCCCCCCCCGGTTATATCCAACAGATCCATATCCACAAGGACTGGGA 2159

Qy    2391 ACTGCCAACCAGAGAAGGCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTG 2450
           || ||||  ||||||||  | |||   |||||  |||.|  ||  |||  |||||||  ||
Db    2160 ACAGCCAGACAGAGAAGAAACAGACGAAGACGCCGGAAACAGCATTGGAGACAGCTCGTG 2219

Qy    2451 GCCTTGGCAGATAGAATATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTG 2510
           ||||||||  |||| ||||||  |||||||||||||||||  ||||| ||||||||||  |
Db    2220 GCCTTGGCCAATAGCATATATACATTTCCTGATCCGCCAGCTGATTCGCCTCTTGACCGG 2279

Qy    2511 GCTATTCAGCAACTGCAGAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACT 2570
           |||||  ||||  |||||    ||| ||||| ||    | || || |||||||| |||  |
Db    2280 GCTATACAGCGTCTGCAAGGACTTACTATCCAGGAGCTTCCCGACCCTCCAACTAATCTT 2339

Qy    2571 CCAGAGGCTCTCTGCGACCCTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTA 2630
           ||||||  ||        | |||   |  ||  | ||    || ||   |  | |  ||||
Db    2340 CCAGAGTCTTCAGAGAGCACTAACAACAATCAGGGACTGGCTGAGACTTACAATAGCCTA 2399

Qy    2631 CCTACAATATG 2641
           |||  |||||||
Db    2400 CCTGCAATATG 2410


RESULT 10
US-09-256-490-13
; Sequence 13, Application US/09256490
; Patent No. 6235881
;  GENERAL INFORMATION:
;    APPLICANT:  Kraus, Guenter
;    APPLICANT:  Wong-Staal, Flossie
;    APPLICANT:  Talbott, Randy
;    APPLICANT:  Poeschla, Eric
;    TITLE OF INVENTION:  Isolation of No. 6235881el HIV-2 Proviruses
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew LLP
;      STREET:  Two Embarcadero Center, Eighth Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/256,490
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/659,251
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Garrett-Wackowski, Eugenia
;      REGISTRATION NUMBER:  37,330
;      REFERENCE/DOCKET NUMBER:  02307E-056410US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 576-0200
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2574 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    FEATURE:
;      NAME/KEY:  -
;      LOCATION:  1..2574
;      OTHER INFORMATION:  /label= env
;      OTHER INFORMATION:  /note= "HIV-2KR subsequence encoding
;      OTHER INFORMATION:  env gene"
US-09-256-490-13

  Query Match             48.0%;  Score 1266.6;  DB 4;  Length 2574;
  Best Local Similarity   72.3%;  Pred. No. 0;
  Matches 1787;  Conservative    0;  Mismatches  609;  Indels   75;  Gaps    8;

Qy     177 TGGGAATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCT 236
           | | ||||||||  || | ||||| || ||    ||||  |      || |||||  | | |
Db       9 TAGAAATCAGCTAATTGTTGCCATTTTACTAACTAGTGCTTGCTTAATATATTGCGCCCA 68

Qy     237 ATATGTCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTG 296
           ||||||  || || || |||||  |||| || |||| ||||||  | ||||||||||||||
Db      69 ATATGTGACTGTTTTCTATGGCATACCCGCGTGGAAGAATGCATCCATTCCCCTCTTTTG 128

Qy     297 TGCAACCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTA 356
           ||||||||  ||||| ||||||||||||||| |  |||||| | ||||| |||| ||||||
Db     129 TGCAACCAGAAATAGAGATACTTGGGGAACCATACAGTGCTTGCCAGACAATGATGATTA 188

Qy     357 TTCAGAAGTGGCCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGA 416
           |   ||| |  |  | ||||| |||||    ||||| || ||||| |||||||| |||||
Db     189 TCAGGAAATACCTTTAAATGTGACAGAGGCTTTTGACGCATGGAACAATACAGTAACAGA 248

Qy     417 ACAGGCAATAGAGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATT 476
           ||| ||| |||| ||||| ||| | || |||||||| ||| |||| || ||||| |||||
Db     249 ACAAGCAGTAGAAGATGTCTGGAATCTATTTGAGACATCAGTAAAACCATGTGTCAAATT 308

Qy     477 ATCCCCATTATGCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGAC 536
           | | || |||||  |    |||  ||| || |  |    ||||
Db     309 AACACCCTTATGTGTGCAAATGGAATGTAACAGCACAAGTACAG---------------- 352

Qy     537 AAAATCAATAACAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACAT 596
                         | ||||  |||| ||  |  ||  | | || ||     ||| ||
Db     353 ----------------AGAGCAGTAACAGCACAAGTGAGGGGAGCA-CAGTCCCAGAGAT 395

Qy     597 GGTCAATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCA 656
             | || || |||| ||| || ||| || |  | |  ||| |||  || |     || |
Db     396 ATTAAACGAAACTACTTCATGCATAACCAACAACAGCTGCTCAGATTTAGGGAGTGAAGA 455

Qy     657 AATGATAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAA 716
             |  |    |||  ||||| ||||||||  || || ||| |||  |  |  | || |
Db     456 GGTAGTCGATTGTCGGTTCAATATGACAGGACTACAACTAGATAAGCCACAGCAATATAG 515

Qy     717 TGAAACTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAG 776
           ||||||  ||||||||     ||| | || |||||        |  |  | ||| |  |  |
Db     516 TGAAACATGGTACTCAAAGGATGTAGTTTGTGA------CACAACTAATGGGACCAGCCG 569

Qy     777 TAGATGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTA 836
             | |||||||||||||| || ||||| || || |||  ||||||  |||||  || || ||
Db     570 CAAATGTTACATGAACCATTGCAACACATCAGTCATCACAGAGTCATGTGATAAGCACTA 629

Qy     837 TTGGGATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAA 896
           ||||||||||| || ||||| || |||||||| || ||||     |||||  |||||  ||
Db     630 TTGGGATGCTATGAGGTTTAGATACTGTGCACCACCGGGTTTATGCTTGCTAAGATGCAA 689

Qy     897 TGACACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCAC 956
           ||| || ||||||||||||||  ||| || ||| |||| || || |   || |||||||
Db     690 TGATACCAATTATTCAGGCTTTGAGCCCAAGTGTCCTAAAGTAGTAGCTGCTACATGCAC 749

Qy     957 AAGGATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAA 1016
           ||| |||||||| || || ||||||||||||||||||||||||||| ||||| ||||||||
Db     750 AAGAATGATGGAAACGCAAACTTCTACTTGGTTTGGCTTTAATGGCACTAGGGCAGAAAA 809

Qy    1017 TAGAACTTATATTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTA 1076
```

```
Db    810 TAGAACATATATCTATTGGCATGGTAGAGATAATAGGACTATTATCAGCTTAAATACACA 869

Qy   1077 TTATAATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCAT 1136
Db    870 TTATAATCTCACAATGCATTGTAAGAGGCCAGGAAATAAGTCAGTTTTGCCAATAACACT 929

Qy   1137 TATGTCTGGATTGGTTTTCCACTCACAACC---AATCAATGATAGGCCAAAGCAGGCATG 1193
Db    930 TAGGTCAGGGAGAGTGTTTCACTCCCGACCGATCATCAATGAAAGACCCAAGCAGGCATG 989

Qy   1194 GTGTTGGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAA 1253
Db    990 GTGCTGGTTCGGAGGTGATTGGAAGAAAGCCATGCAGGAGGTGAAACAAACCCTTGTGAA 1049

Qy   1254 ACATCCCAGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGAGG 1313
Db   1050 ACATCCCAGGTATAGAGGAACCAACGACACACAGAAAATTAACTTTACACAACCAGGAAA 1109

Qy   1314 AGG---AGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTG 1370
Db   1110 AGGTTCAGATGCAGAAGTGGTATACATGTGGACTAACTGCAGAGGAGAATTTCTATACTG 1169

Qy   1371 TAAAATGAATTGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAA 1430
Db   1170 CAACATGACTCGGTTCCTCAATTGGATAGAAAACAG---------------------- 1205

Qy   1431 GGAACAGCATAAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCA 1490
Db   1206 GGCACACCCACAGCGCAATTATGCACCGTGCCATATAAGGCAAATAATTAATACCTGGCA 1265

Qy   1491 TAAAGTAGGCAAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCAC 1550
Db   1266 TAGAGTAGGCCAAAATATATATTTGCCTCCTAGGGAAGGGGAATTGGTCTGCAACTCAAC 1325

Qy   1551 AGTGACCAGTCTCATAGCAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCAT 1610
Db   1326 AGTAACCAGCATAATTGCTAACATTGACATGTTTGAT---AACCAGACAAGCATTACCTT 1382

Qy   1611 GAGTGCAGAGGTGGCAGAACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGAT 1670
Db   1383 TAGTGCAGAGGTGGCAGAACTATACCGATTGGAATTGGGAGATTACAAATTAGTAGAAAT 1442

Qy   1671 CACTCCAATTGGCTTGGCCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAG 1730
Db   1443 AACACCAATTGGCTTCGCACCTACATCAGAAAAAAGGTATTCCTCTGCT---CCACAGAG 1499

Qy   1731 AAATAAAAGAGGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGC 1790
Db   1500 GAATAAAAGAGGTGTGTTTGTGCTAGGAGTCTTGGGTTTTCTCGCAACAGCAGGTTCTGC 1559

Qy   1791 AATGGGCGCGGCGTCGTTGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGT 1850
Db   1560 AATGGGCGCGGCGTCCTTGACGCTGTCGGCTCATCCCGGACTTTACTGGGCTGGGATAGT 1619

Qy   1851 GCAGCAACAGCAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGAC 1910
Db   1620 GCAGCAACAGCAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAAATGTTGCGACTGAC 1679

Qy   1911 CGTCTGGGGAACAAAGAACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGA 1970
Db   1680 CGTCTGGGGAACAAAAAATCTCCAGACAAGAGTCACTGCTATCGAGAAATACCTAAGGGA 1739

Qy   1971 CCAGGCGCAGCTGAATGCTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACC 2030
Db   1740 CCAGGCGCGGCTAAATTCATGGGGATGTGCATTTAGACAAGTCTGCTACACCACTGTACT 1799

Qy   2031 ATGGCCAAATGCAAGTCTAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCG 2090
Db   1800 ATGGGAAAATAACAGCATAGTACCTGATTGGAACAACATGACGTGGCAGGAATGGGAACA 1859

Qy   2091 AAAGGTTGACTTCTTGGAAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACA 2150
```

```
Db   1860 ACAAACCCGCGACCTAGAGGCAAATATCAGTAGATCGTTAGAGCAGGCACAAATCCAACA 1919

Qy   2151 AGAGAAGAACATGTATGAATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTT 2210
Db   1920 AGAGAAAAATATGTATGAGCTACAAAAATTAAATAGCTGGGATGTTTTTGGCAACTGGTT 1979

Qy   2211 TGACCTTGCTTCTTGGATAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAAT 2270
Db   1980 TGATTTAACCTCCTGGATTAAGTATATTCAGTATGGAGTTTATGTAATAATAGGAATAAT 2039

Qy   2271 ACTGTTAAGAATAGTGATCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAG 2330
Db   2040 AGCTTTAAGAATAGTAATATATGTAGTACAATTACTAAGTAGACTTAGAAAGGGCTATAG 2099

Qy   2331 GCCAGTGTTCTCTTCCCCACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGC 2390
Db   2100 GCCTGTTTTCTCTTCCCCCCCCGGTTATATCCAACAGATCCATATCCACAAGGACTGGGA 2159

Qy   2391 ACTGCCAACCAGAGAAGGCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTG 2450
Db   2160 ACAGCCAGACAGAGAAGAAACAGACGAAGACGCCGGAAACAGCATTGGAGACAGCTCGTG 2219

Qy   2451 GCCTTGGCAGATAGAATATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTG 2510
Db   2220 GCCTTGGCCAATAGCATATATACATTTCCTGATCCGCCAGCTGATTCGCCTCTTGACCGG 2279

Qy   2511 GCTATTCAGCAACTGCAGAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACT 2570
Db   2280 GCTATACAGCGTCTGCAAGGACTTACTATCCAGGAGCTTCCCGACCCTCCAACTAATCTT 2339

Qy   2571 CCAGAGGCTCTCTGCGACCCTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTA 2630
Db   2340 CCAGAGTCTTCAGAGAGCACTAACAACAATCAGGGACTGGCTGAGACTTACAATAGCCTA 2399

Qy   2631 CCTACAATATG 2641
Db   2400 CCTGCAATATG 2410
```

```
RESULT 11
PCT-US96-11445-13
; Sequence 13, Application PC/TUS9611445
;  GENERAL INFORMATION:
;    APPLICANT:  The Regents of the University of California
;    TITLE OF INVENTION:  Isolation of Novel HIV-2 Proviruses
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Robbins, Berliner & Carson
;      STREET:  201 N. Figueroa Street, 5th Floor
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  90012-2628
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/11445
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berliner, Robert
;      REGISTRATION NUMBER:  20,121
;      REFERENCE/DOCKET NUMBER:  5555-399C1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 977-1001
;      TELEFAX:  (213) 977-1003
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2574 base pairs
;      TYPE:  nucleic acid
```

```
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    FEATURE:
;      NAME/KEY:  -
;      LOCATION:  1..2574
;      OTHER INFORMATION:  /label= env
;      OTHER INFORMATION:  /note= "HIV-2KR subsequence encoding
;      OTHER INFORMATION:  env gene"
PCT-US96-11445-13

  Query Match             48.0%;  Score 1266.6;  DB 5;  Length 2574;
  Best Local Similarity   72.3%;  Pred. No. 0;
  Matches 1787;  Conservative    0;  Mismatches  609;  Indels   75;  Gaps    8;

Qy    177 TGGGAATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATCTATTGTACTCT 236
          | | ||||||||  || | ||||| || ||    ||||  |     || |||||  | |
Db      9 TAGAAATCAGCTAATTGTTGCCATTTTACTAACTAGTGCTTGCTTAATATATTGCGCCCA 68

Qy    237 ATATGTCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATTCCCCTCTTTTG 296
          ||||||  || || ||  |||||  |||| || |||| ||||||  | ||||||||||||||
Db     69 ATATGTGACTGTTTTCTATGGCATACCCGCGTGGAAGAATGCATCCATTCCCCTCTTTTG 128

Qy    297 TGCAACCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTA 356
          ||||||||  |||||  |||||||||||||||  |  |||||| | |||||  |||| ||||||
Db    129 TGCAACCAGAAATAGAGATACTTGGGGAACCATACAGTGCTTGCCAGACAATGATGATTA 188

Qy    357 TTCAGAAGTGGCCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGA 416
          |   ||| |  |  | |||||  |||||    ||||| || |||||  ||||||||  |||||
Db    189 TCAGGAAATACCTTTAAATGTGACAGAGGCTTTTGACGCATGGAACAATACAGTAACAGA 248

Qy    417 ACAGGCAATAGAGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATT 476
          ||| ||| |||| |||||  ||| |  || ||||||||  ||| |||| ||  |||||  |||||
Db    249 ACAAGCAGTAGAAGATGTCTGGAATCTATTTGAGACATCAGTAAAACCATGTGTCAAATT 308

Qy    477 ATCCCCATTATGCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGAC 536
          | | ||  |||||  |    |||  ||| || |  |      ||||
Db    309 AACACCCTTATGTGTGCAAATGGAATGTAACAGCACAAGTACAG---------------- 352

Qy    537 AAAATCAATAACAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACAT 596
                    | ||||  |||| ||  |  ||  | | || ||     ||| ||
Db    353 ----------------AGAGCAGTAACAGCACAAGTGAGGGGAGCA-CAGTCCCAGAGAT 395

Qy    597 GGTCAATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCA 656
            | || || |||| ||| || ||| || |  | |  ||| |||  || |      ||  |
Db    396 ATTAAACGAAACTACTTCATGCATAACCAACAACAGCTGCTCAGATTTAGGGAGTGAAGA 455

Qy    657 AATGATAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAA 716
            |   |||   ||||| ||||||||  || ||  ||| |||  | |  | || |
Db    456 GGTAGTCGATTGTCGGTTCAATATGACAGGACTACAACTAGATAAGCCACAGCAATATAG 515

Qy    717 TGAAACTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAG 776
          ||||||  ||||||||     ||| | || |||||        |  |  | ||| |       |
Db    516 TGAAACATGGTACTCAAAGGATGTAGTTTGTGA------CACAACTAATGGGACCAGCCG 569

Qy    777 TAGATGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTA 836
            | ||||||||||||||| || ||||| || || |||  |||||| ||||| || || ||
Db    570 CAAATGTTACATGAACCATTGCAACACATCAGTCATCACAGAGTCATGTGATAAGCACTA 629

Qy    837 TTGGGATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAA 896
          ||||||||||||  || ||||| || |||||||| || ||||     |||||  ||||| ||
Db    630 TTGGGATGCTATGAGGTTTAGATACTGTGCACCACCGGGTTTATGCTTGCTAAGATGCAA 689

Qy    897 TGACACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCAC 956
          ||| || ||||||||||||||||  ||| || ||| |||| || || |   || |||||||
Db    690 TGATACCAATTATTCAGGCTTTGAGCCCAAGTGTCCTAAAGTAGTAGCTGCTACATGCAC 749

Qy    957 AAGGATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGCAGAAAA 1016
          ||| |||||||| || || ||||||||||||||||||||||||||| ||||| ||||||||
Db    750 AAGAATGATGGAAACGCAAACTTCTACTTGGTTTGGCTTTAATGGCACTAGGGCAGAAAA 809

Qy   1017 TAGAACTTATATTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTA 1076
          ||||||  |||||  || |||||||||||  ||||||||||||||  || || |||||||   |

Db    810 TAGAACATATATCTATTGGCATGGTAGAGATAATAGGACTATTATCAGCTTAAATACACA 869

Qy   1077 TTATAATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCAT 1136
          |||||||||  ||||||  | ||||  ||  ||||||||||||  |||||||  |||  | ||  |
Db    870 TTATAATCTCACAATGCATTGTAAGAGGCCAGGAAATAAGTCAGTTTTGCCAATAACACT 929

Qy   1137 TATGTCTGGATTGGTTTTCCACTCACAACC---AATCAATGATAGGCCAAAGCAGGCATG 1193
          || ||| ||    || || |||||  | |||    ||||||||  || ||  ||||||||||
Db    930 TAGGTCAGGGAGAGTGTTTCACTCCCGACCGATCATCAATGAAAGACCCAAGCAGGCATG 989

Qy   1194 GTGTTGGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAA 1253
          ||| |||||  |||||  | ||||||  | || ||  | ||||||||  || ||| |||| ||
Db    990 GTGCTGGTTCGGAGGTGATTGGAAGAAAGCCATGCAGGAGGTGAAACAAACCCTTGTGAA 1049

Qy   1254 ACATCCCAGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGAGG 1313
          ||||||||||||||  |||||  ||| |  ||  | |||||  ||  ||  ||    ||  |||
Db    1050 ACATCCCAGGTATAGAGGAACCAACGACACACAGAAAATTAACTTTACACAACCAGGAAA 1109

Qy   1314 AGG---AGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTG 1370
          |||    ||||  |  |||||     | ||||||||  ||  ||||||||||  ||  ||  |||||
Db    1110 AGGTTCAGATGCAGAAGTGGTATACATGTGGACTAACTGCAGAGGAGAATTTCTATACTG 1169

Qy   1371 TAAAATGAATTGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAA 1430
           ||  |||| |  |||| ||  |||||| |||||  | ||
Db    1170 CAACATGACTCGGTTCCTCAATTGGATAGAAAACAG---------------------- 1205

Qy   1431 GGAACAGCATAAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCA 1490
          ||  |||  |    |  | |||||  |  ||  ||  |||||  ||  ||||||||  ||  ||  |||||
Db    1206 GGCACACCCACAGCGCAATTATGCACCGTGCCATATAAGGCAAATAATTAATACCTGGCA 1265

Qy   1491 TAAAGTAGGCAAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCAC 1550
          ||  |||||||  |||||  |  ||||||||||  ||  ||  ||  ||  |    ||  |||||  ||
Db    1266 TAGAGTAGGCCAAAATATATATTTGCCTCCTAGGGAAGGGGAATTGGTCTGCAACTCAAC 1325

Qy   1551 AGTGACCAGTCTCATAGCAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCAT 1610
          |||  |||||  |  ||  ||  |||||  ||    |  |||||    |||||  ||  |  ||  |||  |
Db    1326 AGTAACCAGCATAATTGCTAACATTGACATGTTTGAT---AACCAGACAAGCATTACCTT 1382

Qy   1611 GAGTGCAGAGGTGGCAGAACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAGAT 1670
          ||||||||||||||||||||  ||  ||||||||||||||||||||  ||||||||||  ||
Db    1383 TAGTGCAGAGGTGGCAGAACTATACCGATTGGAATTGGGAGATTACAAATTAGTAGAAAT 1442

Qy   1671 CACTCCAATTGGCTTGGCCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCAAG 1730
          ||  ||||||||||  ||  ||  |||    |  ||  |||||  |  |||  |    |  ||
Db    1443 AACACCAATTGGCTTCGCACCTACATCAGAAAAAAGGTATTCCTCTGCT---CCACAGAG 1499

Qy   1731 AAATAAAAGAGGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTCTGC 1790
          ||||||||||  ||  ||||||||||  ||||||||||||||||||||  ||||||||||
Db    1500 GAATAAAAGAGGTGTGTTTGTGCTAGGAGTCTTGGGTTTTCTCGCAACAGCAGGTTCTGC 1559

Qy   1791 AATGGGCGCGGCGTCGTTGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGATAGT 1850
          ||||||||||||||| |||||||| | |||||  ||  ||  ||  | ||||||||||||
Db    1560 AATGGGCGCGGCGTCCTTGACGCTGTCGGCTCATCCCGGACTTTACTGGGCTGGGATAGT 1619

Qy   1851 GCAGCAACAGCAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGAC 1910
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1620 GCAGCAACAGCAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAAATGTTGCGACTGAC 1679

Qy   1911 CGTCTGGGGAACAAAGAACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAAGGA 1970
          ||||||||||||||| || |||||||| || |||||||  |||||||| ||| ||| |||
Db    1680 CGTCTGGGGAACAAAAAATCTCCAGACAAGAGTCACTGCTATCGAGAAATACCTAAGGGA 1739

Qy   1971 CCAGGCGCAGCTGAATGCTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACC 2030
          ||||||||  ||| ||| | ||||||||||  |||||||||||||  |||| |||||||
Db    1740 CCAGGCGCGGCTAAATTCATGGGGATGTGCATTTAGACAAGTCTGCTACACCACTGTACT 1799

Qy   2031 ATGGCCAAATGCAAGTCTAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCG 2090
          ||||  ||||   ||  ||  |||  | ||||||||   ||| |||||  || |||| |
Db    1800 ATGGGAAAATAACAGCATAGTACCTGATTGGAACAACATGACGTGGCAGGAATGGGAACA 1859

Qy   2091 AAAGGTTGACTTCTTGGAAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATTCAACA 2150
          | |   |  | | || | ||||||  |          |||||  ||||||||||  |||||
Db    1860 ACAAACCCGCGACCTAGAGGCAAATATCAGTAGATCGTTAGAGCAGGCACAAATCCAACA 1919
```

```
Qy   2151 AGAGAAGAACATGTATGAATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTT 2210
Db   1920 AGAGAAAAATATGTATGAGCTACAAAAATTAAATAGCTGGGATGTTTTTGGCAACTGGTT 1979

Qy   2211 TGACCTTGCTTCTTGGATAAAGTATATACAATATGGAGTTTATATAGTTGTAGGAGTAAT 2270
Db   1980 TGATTTAACCTCCTGGATTAAGTATATTCAGTATGGAGTTTATGTAATAATAGGAATAAT 2039

Qy   2271 ACTGTTAAGAATAGTGATCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAG 2330
Db   2040 AGCTTTAAGAATAGTAATATATGTAGTACAATTACTAAGTAGACTTAGAAAGGGCTATAG 2099

Qy   2331 GCCAGTGTTCTCTTCCCCACCCTCTTATTTCCAGCAGACCCATATCCAACAGGACCCGGC 2390
Db   2100 GCCTGTTTTCTCTTCCCCCCCCGGTTATATCCAACAGATCCATATCCACAAGGACTGGGA 2159

Qy   2391 ACTGCCAACCAGAGAAGGCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTG 2450
Db   2160 ACAGCCAGACAGAGAAGAAACAGACGAAGACGCCGGAAACAGCATTGGAGACAGCTCGTG 2219

Qy   2451 GCCTTGGCAGATAGAATATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTG 2510
Db   2220 GCCTTGGCCAATAGCATATATACATTTCCTGATCCGCCAGCTGATTCGCCTCTTGACCGG 2279

Qy   2511 GCTATTCAGCAACTGCAGAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCAATACT 2570
Db   2280 GCTATACAGCGTCTGCAAGGACTTACTATCCAGGAGCTTCCCGACCCTCCAACTAATCTT 2339

Qy   2571 CCAGAGGCTCTCTGCGACCCTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTGACCTA 2630
Db   2340 CCAGAGTCTTCAGAGAGCACTAACAACAATCAGGGACTGGCTGAGACTTACAATAGCCTA 2399

Qy   2631 CCTACAATATG 2641
Db   2400 CCTGCAATATG 2410


RESULT 12
US-08-944-449-8
; Sequence 8, Application US/08944449
; Patent No. 5985613
; GENERAL INFORMATION:
;  APPLICANT: KURTH, REINHARD
;  APPLICANT: BAIER, MICHAEL
;  APPLICANT: METZNER, KARIN
;  APPLICANT: WERNER, ALBRECHT
;  TITLE OF INVENTION: Use of an "immunodeficiency-virus suppressing
;  TITLE OF INVENTION: lymphokine (ISL)" to inhibit the replication of
;  TITLE OF INVENTION: viruses, particularly of retroviruses
;  FILE REFERENCE: 8341-7065
;  CURRENT APPLICATION NUMBER: US/08/944,449
;  CURRENT FILING DATE: 1997-10-06
;  EARLIER APPLICATION NUMBER: EP 95113013.2
;  EARLIER FILING DATE: 1995-08-18
;  EARLIER APPLICATION NUMBER: DE 195 13 152.5
;  EARLIER FILING DATE: 1995-04-07
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
;   LENGTH: 4527
;   TYPE: DNA
;   ORGANISM: Human immunodeficiency virus type 1
US-08-944-449-8

  Query Match             9.5%;  Score 251.6;  DB 2;  Length 4527;
  Best Local Similarity   49.9%;  Pred. No. 4.8e-59;
  Matches  969;  Conservative    0;  Mismatches  914;  Indels   59;  Gaps   11;

Qy    604 GAGACTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCAAATGATA 663
Db    884 GGGAATGCTACTAATACCAACAATAGTAGTGGGGGAACGGTGGAGAAAGAAGAAATAAAA 943

Qy    664 AGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAATGAAACT 723
Db    944 AACTGCTCTTTCAATATCACCACAGGCATAAGAGATAAGGTACAGAAGGCATATGCATAT 1003

Qy    724 TGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGATGT 783
Db   1004 TTTTATAAACTTGATGTAGTACCAATAGATGATGATAATACTAATACCAGCTATAGGT-- 1061

Qy    784 TACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGGGAT 843
Db   1062 ----TGATACATTGTAATTCCTCAGTCATTACACAGACCTGTCCAAAGGTATCCTTTGAG 1117

Qy    844 GCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGACACA 903
Db   1118 CCAATTCCTATACATTATTGTGCCCCGGCTGGTTTTGCGATTCTAAAGTGTAATAATAAG 1177

Qy    904 AATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGGATG 963
Db   1178 AAGTTCAGTGG---AAAAGGTCAATGTACAAATGTCAGCACAGTACAATGTACACATGGA 1234

Qy    964 ATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGC---AGAAAATAGA 1020
Db   1235 ATTAAGCCAGTAGTGTCAACTCAACTGCTGTTAAATGGCAGTCTAGCAGAAGAAGAGGTA 1294

Qy   1021 ACTTATATTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTAT 1080
Db   1295 GTAATTAGATCTGACAATTTCACGAACAATGCTAAAACCATATTAGTACAGCTGAATGTA 1354

Qy   1081 AATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATG 1140
Db   1355 TCTGTAGAAATTAATTGTACAAGACCCAACAACAATAGAAGAAGAAGGATAACTAGTGGA 1414

Qy   1141 TCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGGCCAAAGCAGGCATGGTGTTGG 1200
Db   1415 CCAGGGAAAGTACTTTA--TACAACAGGAGAAATAATAGGAGATATAAGAAAAGCATATT 1472

Qy   1201 TTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCC 1260
Db   1473 GTAACATTAGTAGAGCAAAATGGAATAAAACTTTAGAACAGGTAGCTACAAAATTAAGAG 1532

Qy   1261 AGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGAGGAGGAGAT 1320
Db   1533 AACAATTTGGGAATAAAACAATAGTATTTAAACAAT----------CCTCAGGAGGAGAC 1582

Qy   1321 CCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATGAAT 1380
Db   1583 CCAGAAATTGTAATGCACAGTTTTAATTGTAGAGGGGAATTTTTCTACTGTAATACAACA 1642

Qy   1381 TGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGGAACAGCAT 1440
Db   1643 AAACTGTTTAAT--AGTACTTGGAATGAAAATAGTACTTGGAATGCTACTGGAA----AT 1696

Qy   1441 AAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAGGC 1500
Db   1697 GACACTATCACACTCCCATGTAGAATAAAACAAATTATAAACATGTGGCAGGAAGTAGGA 1756

Qy   1501 AAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACCAGT 1560
Db   1757 AAAGCAATGTATGCCCCTCCCATCGAAGGACAAATTAGATGTTCATCAAATATTACAGGG 1816

Qy   1561 CTCATAGCAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATG--------- 1611
Db   1817 CTGCTATTAACAAGAGATGGTGGTGGTGACAAGAACAGTACCACCGAGATCTTTAGACCT 1876

Qy   1612 ---AGTGCAGAGGTGGCAGAACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAG 1668
Db   1877 GCAGGAGGAAATATGAAGGACAATTGGAGAAGTGAATTATATAAATATAAAGTAGTAAAA 1936

Qy   1669 ATCACTCCAATTGGCTTGGCCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCA 1728
Db   1937 ATTGAACCATTAGGAGTAGCACCCACCAAGGCAAAGAG----AAGAGTGGTGCAAAGAGA 1992

Qy   1729 AGAAATAAAAGAGGG-GTCTTTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTC 1787
```

```
Db    1993 AAAAAGAGCAGTGGGAGTGATAGGAGCTATGTTCCTTGGGTTCTTGGGAGCAGCAGGAAG 2052

Qy    1788 TGCAATGGGCGCGGCGTCGTTGACGCTGACCGCTGAGTCCCGAACTTTATTGGCTGGGAT 1847

Db    2053 CACTATGGGCGCAGCGTCAATAACGCTGACGGTACAGGCCAGAAAACTATTGTCTGGTAT 2112

Qy    1848 AGTGCAGCAACAGCAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACT 1907

Db    2113 AGTGCAACAGCAGAACAATCTGCTGAGAGCTATTGAGGCGCAACAGCATCTGTTGCAACT 2172

Qy    1908 GACCGTCTGGGGAACAAAGAACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAA 1967

Db    2173 CACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAGTCCTGGCTGTGGAAAGATACCTAAG 2232

Qy    1968 GGACCAGGCGCAGCTGAATGCTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGT 2027

Db    2233 AGATCAACAGCTCCTAGGGATTTGGGGTTGCTCTGGAAAACTCATTTGCACCACTACTGT 2292

Qy    2028 ACC------------ATGGCCAAATGCAAGTCTAACACCAAAGTGGAACAATGAGACTTG 2075

Db    2293 GCCTTGGAATACTAGTTGGAGTAATAAATCTCTGGATAAGATTTGGAATAACATGACTTG 2352

Qy    2076 GCAAGAGTGGGAGCGAAAGGTTGACTTCTTGGAAGAAAATATAACAGCCCTCCTAGAGGA 2135

Db    2353 GATGGAGTGGGAAAGAGAAATTGACAATTACACAAGCTTAATATACACCTTACTTGAAGA 2412

Qy    2136 GGCACAAATTCAACAAGAGAAGAACATGTATGAATTACAAAAGTTGAATAGCTGGGATGT 2195

Db    2413 ATCGCAAAACCAACAAGAAAAGAATGAACAAGAGTTATTGGAATTGGATAAGTGGGCAAG 2472

Qy    2196 GTTTGGCAATTGGTTTGACCTTGCTTCTTGGATAAAGTATATACAATATGGAGTTTATAT 2255

Db    2473 TTTGTGGAATTGGTTTAGCATAACAAACTGGCTGTGGTATATAAGAATATTCATAATGAT 2532

Qy    2256 AGTTGTAGGAGTAATACTGTTAAGAATAGTGATCTATATAGTACAAATGCTAGCTAAGTT 2315

Db    2533 AGTAGGAGGCTTGATAGGTTTAAGAATAATTTTTGCTGTGCTTTCTATAGTAAATAGAGT 2592

Qy    2316 AAGGCAGGGGTATAGGCCAGTGTTCTCTTCCCCACCCTCTTATTTCCAGCAGACCCATAT 2375

Db    2593 TAGGCAGGGATACTCACCATTATCATTTCAGACCCTCATCCCAGCCCAGAGGGGACCCGA 2652

Qy    2376 CCAACAGGACCCGGCACTGCCAACCAGAGAAGGCAAAGAAAGAGACGGTGGAGAAGGCGG 2435

Db    2653 CAGGCCCGAAGGAATCGAAGAAGGAGGTGGAGAGAGAGACAGAGACAGATCCACTCGATT 2712

Qy    2436 TGGCAACAGCTCCTGGCCTTGGCAGATAGAATATATTCATTTCCTGATCCGCCAACTGAT 2495

Db    2713 AGTGAACGGATTCTTAGCACTGTTCTGGGACGATCTTCGGAGCCTGTGCCTCTTCAGCTA 2772

Qy    2496 ACGCCTCTTGACTTGGCTATTC 2517

Db    2773 CCACCGCTTGACAGACTTACTC 2794


RESULT 13
US-09-353-362-8
; Sequence 8, Application US/09353362
; Patent No. 6383739
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Use of an "immunodeficiency-virus suppressing
;    TITLE OF INVENTION:  lymphokine (ISL)" to inhibit the replication of viruses,
;    TITLE OF INVENTION:  in particular of retroviruses
;    NUMBER OF SEQUENCES:  8
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/353,362
;      FILING DATE:  15-JUL-1999
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE 195 13 152.5
;      FILING DATE:  07-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 95113013.2
;      FILING DATE:  18-AUG-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KLESNER, Sharon N.
;      REGISTRATION NUMBER:  36,335
;      REFERENCE/DOCKET NUMBER:  P8341-9012
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 638-5000
;      TELEFAX:  (202) 638-4810
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4527 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-353-362-8

  Query Match             9.5%;  Score 251.6;  DB 4;  Length 4527;
  Best Local Similarity   49.9%;  Pred. No. 4.8e-59;
  Matches  969;  Conservative    0;  Mismatches  914;  Indels   59;  Gaps   11;

Qy     604 GAGACTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCAAATGATA 663

Db     884 GGGAATGCTACTAATACCAACAATAGTAGTGGGGGAACGGTGGAGAAAGAAGAAATAAAA 943

Qy     664 AGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAATGAAACT 723

Db     944 AACTGCTCTTTCAATATCACCACAGGCATAAGAGATAAGGTACAGAAGGCATATGCATAT 1003

Qy     724 TGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGATGT 783

Db    1004 TTTTATAAACTTGATGTAGTACCAATAGATGATGATAATACTAATACCAGCTATAGGT-- 1061

Qy     784 TACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAAACATTATTGGGAT 843

Db    1062 ----TGATACATTGTAATTCCTCAGTCATTACACAGACCTGTCCAAAGGTATCCTTTGAG 1117

Qy     844 GCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAGATGTAATGACACA 903

Db    1118 CCAATTCCTATACATTATTGTGCCCCGGCTGGTTTTGCGATTCTAAAGTGTAATAATAAG 1177

Qy     904 AATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTCATGCACAAGGATG 963

Db    1178 AAGTTCAGTGG---AAAAGGTCAATGTACAAATGTCAGCACAGTACAATGTACACATGGA 1234

Qy     964 ATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGC---AGAAAATAGA 1020

Db    1235 ATTAAGCCAGTAGTGTCAACTCAACTGCTGTTAAATGGCAGTCTAGCAGAAGAAGAGGTA 1294

Qy    1021 ACTTATATTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTAT 1080

Db    1295 GTAATTAGATCTGACAATTTCACGAACAATGCTAAAACCATATTAGTACAGCTGAATGTA 1354

Qy    1081 AATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATG 1140

Db    1355 TCTGTAGAAATTAATTGTACAAGACCCAACAACAATAGAAGAAGAAGGATAACTAGTGGA 1414

Qy    1141 TCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGGCCAAAGCAGGCATGGTGTTGG 1200

Db    1415 CCAGGGAAAGTACTTTA--TACAACAGGAGAAATAATAGGAGATATAAGAAAAGCATATT 1472

Qy    1201 TTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCC 1260

Db    1473 GTAACATTAGTAGAGCAAAATGGAATAAAACTTTAGAACAGGTAGCTACAAAATTAAGAG 1532
```

```
Qy   1261 AGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGAGGAGGAGAT 1320
Db   1533 AACAATTTGGGAATAAAACAATAGTATTTAAACAAT----------CCTCAGGAGGAGAC 1582

Qy   1321 CCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATGAAT 1380
Db   1583 CCAGAAATTGTAATGCACAGTTTTAATTGTAGAGGGGAATTTTTCTACTGTAATACAACA 1642

Qy   1381 TGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGGAACAGCAT 1440
Db   1643 AAACTGTTTAAT--AGTACTTGGAATGAAAATAGTACTTGGAATGCTACTGGAA----AT 1696

Qy   1441 AAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAGGC 1500
Db   1697 GACACTATCACACTCCCATGTAGAATAAAACAAATTATAAACATGTGGCAGGAAGTAGGA 1756

Qy   1501 AAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTCCACAGTGACCAGT 1560
Db   1757 AAAGCAATGTATGCCCCTCCCATCGAAGGACAAATTAGATGTTCATCAAATATTACAGGG 1816

Qy   1561 CTCATAGCAAACATAGATTGGATTGATGGAAACCAAACTAATATCACCATG--------- 1611
Db   1817 CTGCTATTAACAAGAGATGGTGGTGGTGACAAGAACAGTACCACCGAGATCTTTAGACCT 1876

Qy   1612 ---AGTGCAGAGGTGGCAGAACTGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAG 1668
Db   1877 GCAGGAGGAAATATGAAGGACAATTGGAGAAGTGAATTATATAAATATAAAGTAGTAAAA 1936

Qy   1669 ATCACTCCAATTGGCTTGGCCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACCTCA 1728
Db   1937 ATTGAACCATTAGGAGTAGCACCCACCAAGGCAAAGAG----AAGAGTGGTGCAAAGAGA 1992

Qy   1729 AGAAATAAAAGAGGG-GTCTTTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGTTC 1787
Db   1993 AAAAAGAGCAGTGGGAGTGATAGGAGCTATGTTCCTTGGGTTCTTGGGAGCAGCAGGAAG 2052

Qy   1788 TGCAATGGGCGCGGCGTCGTTGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGGAT 1847
Db   2053 CACTATGGGCGCAGCGTCAATAACGCTGACGGTACAGGCCAGAAAACTATTGTCTGGTAT 2112

Qy   1848 AGTGCAGCAACAGCAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACT 1907
Db   2113 AGTGCAACAGCAGAACAATCTGCTGAGAGCTATTGAGGCGCAACAGCATCTGTTGCAACT 2172

Qy   1908 GACCGTCTGGGGAACAAAGAACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTAAA 1967
Db   2173 CACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAGTCCTGGCTGTGGAAAGATACCTAAG 2232

Qy   1968 GGACCAGGCGCAGCTGAATGCTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGT 2027
Db   2233 AGATCAACAGCTCCTAGGGATTTGGGGTTGCTCTGGAAAACTCATTTGCACCACTACTGT 2292

Qy   2028 ACC------------ATGGCCAAATGCAAGTCTAACACCAAAGTGGAACAATGAGACTTG 2075
Db   2293 GCCTTGGAATACTAGTTGGAGTAATAAATCTCTGGATAAGATTTGGAATAACATGACTTG 2352

Qy   2076 GCAAGAGTGGGAGCGAAAGGTTGACTTCTTGGAAGAAAATATAACAGCCCTCCTAGAGGA 2135
Db   2353 GATGGAGTGGGAAAGAGAAATTGACAATTACACAAGCTTAATATACACCTTACTTGAAGA 2412

Qy   2136 GGCACAAATTCAACAAGAGAAGAACATGTATGAATTACAAAAGTTGAATAGCTGGGATGT 2195
Db   2413 ATCGCAAAACCAACAAGAAAAGAATGAACAAGAGTTATTGGAATTGGATAAGTGGGCAAG 2472

Qy   2196 GTTTGGCAATTGGTTTGACCTTGCTTCTTGGATAAAGTATATACAATATGGAGTTTATAT 2255
Db   2473 TTTGTGGAATTGGTTTAGCATAACAAACTGGCTGTGGTATATAAGAATATTCATAATGAT 2532

Qy   2256 AGTTGTAGGAGTAATACTGTTAAGAATAGTGATCTATATAGTACAAATGCTAGCTAAGTT 2315
Db   2533 AGTAGGAGGCTTGATAGGTTTAAGAATAATTTTTGCTGTGCTTTCTATAGTAAATAGAGT 2592

Qy   2316 AAGGCAGGGGTATAGGCCAGTGTTCTCTTCCCCACCCTCTTATTTCCAGCAGACCCATAT 2375
```

```
Db   2593 TAGGCAGGGATACTCACCATTATCATTTCAGACCCTCATCCCAGCCCAGAGGGGACCCGA 2652

Qy   2376 CCAACAGGACCCGGCACTGCCAACCAGAGAAGGCAAAGAAAGAGACGGTGGAGAAGGCGG 2435
Db   2653 CAGGCCCGAAGGAATCGAAGAAGGAGGTGGAGAGAGAGACAGAGACAGATCCACTCGATT 2712

Qy   2436 TGGCAACAGCTCCTGGCCTTGGCAGATAGAATATATTCATTTCCTGATCCGCCAACTGAT 2495
Db   2713 AGTGAACGGATTCTTAGCACTGTTCTGGGACGATCTTCGGAGCCTGTGCCTCTTCAGCTA 2772

Qy   2496 ACGCCTCTTGACTTGGCTATTC 2517
Db   2773 CCACCGCTTGACAGACTTACTC 2794


RESULT 14
US-08-254-358-3
; Sequence 3, Application US/08254358
; Patent No. 5658785
;  GENERAL INFORMATION:
;    APPLICANT:  Johnson, Philip R.
;    TITLE OF INVENTION:  Adeno-Associated Virus Materials and
;    TITLE OF INVENTION:  Methods
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 S. Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/254,358
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5658785and, Greta E.
;      REGISTRATION NUMBER:  35,302
;      REFERENCE/DOCKET NUMBER:  31975
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 474-6300
;      TELEFAX:  (312) 474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2571 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-254-358-3

  Query Match             9.2%;  Score 242;  DB 1;  Length 2571;
  Best Local Similarity   50.2%;  Pred. No. 1.6e-56;
  Matches  862;  Conservative    0;  Mismatches  790;  Indels   66;  Gaps    8;

Qy    651 AGAGCAAATGATAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGA 710
Db    468 AGGAGAAATGAAAAACTGCTCTTTCAATATCACCACAAGCATAAGAGATAAGATGCAGAA 527

Qy    711 GTACAATGAAACTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAA 770
Db    528 AGAATATGCACTTCTTTATAAACTTGATATAGTATCAATAGATAATGATAGTAC------ 581

Qy    771 TGAAAGTAGATGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAA 830
Db    582 ------CAGCTATAGGTTGATAAGTTGTAATACCTCAGTCATTACACAAGCTTGTCCAAA 635
```

```
Qy    831 ACATTATTGGGATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAG 890
Db    636 GATATCCTTTGAGCCAATTCCCATACACTATTGTGCCCCGGCTGGTTTTGCGATTCTAAA 695

Qy    891 ATGTAATGACACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTC 950
Db    696 ATGTAACGATAAAAAGTTCAGTGGAAAAGGATC---ATGTAAAAATGTCAGCACAGTACA 752

Qy    951 ATGCACAAGGATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGC 1010
Db    753 ATGTACACATGGAATTAGGCCAGTAGTATCAACTCAACTGCTGTTAAATGGCAGTCTAGC 812

Qy   1011 AGAAAATAGAACTTATATTTACT---GGCATGGTAGGGATAATAGGACTATAATTAGTTT 1067
Db    813 AGAAGAAGAGGTAGTAATTAGATCTGAGAATTTCACTGATAATGCTAAAACCATCATAGT 872

Qy   1068 AAATAAGTATTATAATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACC 1127
Db    873 ACATCTGAATGAATCTGTACAAATTAATTGTACAAGACCCAACTACAATAAAAGAAAAAG 932

Qy   1128 AGTCACCATTATGTCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGGCCAAAGCA 1187
Db    933 GATACATATAGGACCAGGGAGAGCATTTTATACA-ACAAAAAATATAATAGGAACTATAA 991

Qy   1188 GGCATGGTGTTGGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCAT 1247
Db    992 GACAAGCACATTGTAACATTAGTAGAGCAAAATGGAATGACACTTTAAGACAGATAGT-- 1049

Qy   1248 TGTCAAACATCCCAGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCC 1307
Db   1050 ---------TAGCAAATTAAAAGAACAATTTAAGAATAAAACAATAGTCTTTAATCAATC 1100

Qy   1308 TGGAGGAGGAGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTA 1367
Db   1101 CTCAGGAGGGGACCCAGAAATTGTAATGCACAGTTTTAATTGTGGAGGGGAATTTTTCTA 1160

Qy   1368 CTGTAAAATGAATTGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCC 1427
Db   1161 CTGTAATACATCACCACTGTTTAATAGTACTTGGAATGGTAATAATACTTGGAATAATAC 1220

Qy   1428 AAAGGAACAGCATAAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCAACACTTG 1487
Db   1221 TACAGGGTCAAATAACAATATCACACTTCAATGCAAAATAAAACAAATTATAAACATGTG 1280

Qy   1488 GCATAAAGTAGGCAAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTC 1547
Db   1281 GCAGGAAGTAGGAAAAGCAATGTATGCCCCTCCCATTGAAGGACAAATTAGATGTTCATC 1340

Qy   1548 CACAGTGACCAGTCTCATAGCAAACATAGATTGGATTGATGGAAACCAAACTAATATCAC 1607
Db   1341 AAATATTACAGGGCTACTATTAACAAGAGATGGTGGTAAGGACACGGACACGAACGACAC 1400

Qy   1608 CATGA---------------GTGCAGAGGTGGCAGAACTGTATCGATTGGAATTGGGAGA 1652
Db   1401 CGAGATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAA 1460

Qy   1653 TTATAAATTAGTAGAGATCACTCCAATTGGCTTGGCCCCCACAGATGTGAAGAGGTACAC 1712
Db   1461 ATATAAAGTAGTAACAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAAGAG------ 1514

Qy   1713 TACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCT 1772
Db   1515 ---AAGAGTGGTGCAGAGAGAAAAAAGAGCAGCGATAGGAGCTCTGTTCCTTGGGTTCTT 1571

Qy   1773 CGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGTTGACGCTGACCGCTCAGTCCCGAAC 1832
Db   1572 AGGAGCAGCAGGAAGCACTATGGGCGCAGCGTCAGTGACGCTGACGGTACAGGCCAGACT 1631

Qy   1833 TTTATTGGCTGGGATAGTGCAGCAACAGCAACAGCTGTTGGACGTGGTCAAGAGACAACA 1892
Db   1632 ATTATTGTCTGGTATAGTGCAACAGCAGAACAATTTGCTGAGGGCCATTGAGGCGCAACA 1691

Qy   1893 AGAATTGTTGCGACTGACCGTCTGGGGAACAAAGAACCTCCAGACTAGGGTCACTGCCAT 1952
Db   1692 GCATATGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAGTCCTGGCTGT 1751

Qy   1953 CGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATGCTTGGGGATGTGCGTTTAGACAAGT 2012
Db   1752 GGAAAGATACCTAAAGGATCAACAGCTCCTGGGGTTTTGGGGTTGCTCTGGAAAACTCAT 1811

Qy   2013 CTGCCACACTACTGTACCATGGCCA-----------AATGCAAGTCTAACACCAAAGTG 2060
Db   1812 TTGCACCACTACTGTGCCTTGGAATGCTAGTTGGAGTAATAAATCTCTGGATGATATTTG 1871

Qy   2061 GAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGGAAGAAAATATAAC 2120
Db   1872 GAATAACATGACCTGGATGCAGTGGGAAAGAGAAATTGACAATTACACAAGCTTAATATA 1931

Qy   2121 AGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATGAATTACAAAAGTT 2180
Db   1932 CTCATTACTAGAAAAATCGCAAACCCAACAAGAAAAGAATGAACAAGAATTATTGGAATT 1991

Qy   2181 GAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGATAAAGTATATACA 2240
Db   1992 GGATAAATGGGCAAGTTTGTGGAATTGGTTTGACATAACAAATTGGCTGTGGTATATAAA 2051

Qy   2241 ATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGATCTATATAGTACA 2300
Db   2052 AATATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGCTGTACTTTC 2111

Qy   2301 AATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGT 2338
Db   2112 TATAGTGAATAGAGTTAGGCAGGGATACTCACCATTGT 2149
```

RESULT 15
US-08-475-391-3
; Sequence 3, Application US/08475391
; Patent No. 5786211
; GENERAL INFORMATION:
; APPLICANT: Johnson, Philip R.
; TITLE OF INVENTION: Adeno-Associated Virus Materials and
; TITLE OF INVENTION: Methods
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 S. Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,391
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/254,358
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5786211and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31975
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2571 base pairs
; TYPE: nucleic acid

```
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-475-391-3

  Query Match             9.2%;  Score 242;  DB 1;  Length 2571;
  Best Local Similarity   50.2%;  Pred. No. 1.6e-56;
  Matches  862;  Conservative    0;  Mismatches  790;  Indels   66;  Gaps    8;

Qy    651 AGAGCAAATGATAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGA 710
          ||   |||||| || |||    ||||| || ||       | |||||| ||||   |  |
Db    468 AGGAGAAATGAAAAACTGCTCTTTCAATATCACCACAAGCATAAGAGATAAGATGCAGAA 527

Qy    711 GTACAATGAAACTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAA 770
            |  ||| |  |   ||       ||| | ||||    | |     |||  ||
Db    528 AGAATATGCACTTCTTTATAAACTTGATATAGTATCAATAGATAATGATAGTAC------ 581

Qy    771 TGAAAGTAGATGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAA 830
              || | |    |||     ||||| || || || ||   | |  |||||   ||
Db    582 ------CAGCTATAGGTTGATAAGTTGTAATACCTCAGTCATTACACAAGCTTGTCCAAA 635

Qy    831 ACATTATTGGGATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAG 890
            |  |  ||  | |||    |    |||||||| ||  | |||| |||  | || |
Db    636 GATATCCTTTGAGCCAATTCCCATACACTATTGTGCCCCGGCTGGTTTTGCGATTCTAAA 695

Qy    891 ATGTAATGACACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTC 950
          |||||| || | ||| |     ||       |   ||||   || ||
Db    696 ATGTAACGATAAAAAGTTCAGTGGAAAAGGATC---ATGTAAAAATGTCAGCACAGTACA 752

Qy    951 ATGCACAAGGATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGC 1010
          ||| |||      ||    | ||      || |||     |     || ||||| | |  |||
Db    753 ATGTACACATGGAATTAGGCCAGTAGTATCAACTCAACTGCTGTTAAATGGCAGTCTAGC 812

Qy   1011 AGAAAATAGAACTTATATTTACT---GGCATGGTAGGGATAATAGGACTATAATTAGTTT 1067
          |||| |          |||   |    | ||   |  ||||||   |  |  || |   |
Db    813 AGAAGAAGAGGTAGTAATTAGATCTGAGAATTTCACTGATAATGCTAAAACCATCATAGT 872

Qy   1068 AAATAAGTATTATAATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACC 1127
          | ||  | || |   | ||  ||| || |||| ||||||     | || | |     |
Db    873 ACATCTGAATGAATCTGTACAAATTAATTGTACAAGACCCAACTACAATAAAAGAAAAAG 932

Qy   1128 AGTCACCATTATGTCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGGCCAAAGCA 1187
            |    ||     | ||    |  ||  |  || |   ||   || |||||    |  |
Db    933 GATACATATAGGACCAGGGAGAGCATTTTATACA-ACAAAAAATATAATAGGAACTATAA 991

Qy   1188 GGCATGGTGTTGGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCAT 1247
          | || |    | ||          |  | ||    | |  | |       ||   | |
Db    992 GACAAGCACATTGTAACATTAGTAGAGCAAAATGGAATGACACTTTAAGACAGATAGT-- 1049

Qy   1248 TGTCAAACATCCCAGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCC 1307
                |  ||  |  |  | |      || | | | | |||    ||  |      |
Db   1050 ---------TAGCAAATTAAAAGAACAATTTAAGAATAAAACAATAGTCTTTAATCAATC 1100

Qy   1308 TGGAGGAGGAGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTA 1367
             |||||| || || ||| ||    |      |    |||||  |||| || ||  ||||
Db   1101 CTCAGGAGGGGACCCAGAAATTGTAATGCACAGTTTTAATTGTGGAGGGGAATTTTTCTA 1160

Qy   1368 CTGTAAAATGAATTGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCC 1427
          ||||||  |          |  | ||| |          || | ||||    ||     | ||  |
Db   1161 CTGTAATACATCACCACTGTTTAATAGTACTTGGAATGGTAATAATACTTGGAATAATAC 1220

Qy   1428 AAAGGAACAGCATAAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCAACACTTG 1487
            |  |      |||| |  |      | | |||  | || | ||||||  || ||||  ||
Db   1221 TACAGGGTCAAATAACAATATCACACTTCAATGCAAAATAAAACAAATTATAAACATGTG 1280

Qy   1488 GCATAAAGTAGGCAAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTC 1547
          |||  |||||||  |||    | |||   |||||  |  || |||  |  |  |  |||   ||
Db   1281 GCAGGAAGTAGGAAAAGCAATGTATGCCCCTCCCATTGAAGGACAAATTAGATGTTCATC 1340

Qy   1548 CACAGTGACCAGTCTCATAGCAAACATAGATTGGATTGATGGAAACCAAACTAATATCAC 1607
            |  | ||  | ||  ||  ||  | |||| |   | | |   |  || ||   |||
Db   1341 AAATATTACAGGGCTACTATTAACAAGAGATGGTGGTAAGGACACGGACACGAACGACAC 1400

Qy   1608 CATGA---------------GTGCAGAGGTGGCAGAACTGTATCGATTGGAATTGGGAGA 1652
            |  ||               | | |||  ||   ||    |   ||   |||||     |
Db   1401 CGAGATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAA 1460

Qy   1653 TTATAAATTAGTAGAGATCACTCCAATTGGCTTGGCCCCCACAGATGTGAAGAGGTACAC 1712
            |||||| |||||   ||    ||| | ||  | || |||||  | |  |||||
Db   1461 ATATAAAGTAGTAACAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAAGAG------ 1514

Qy   1713 TACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCT 1772
                | |         ||| | |||||||   |   | |        |||| | || ||  |
Db   1515 ---AAGAGTGGTGCAGAGAGAAAAAAGAGCAGCGATAGGAGCTCTGTTCCTTGGGTTCTT 1571

Qy   1773 CGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGTTGACGCTGACCGCTCAGTCCCGAAC 1832
            | | | |||||     | ||||||||| |||||  |||||||||| |  ||| || ||
Db   1572 AGGAGCAGCAGGAAGCACTATGGGCGCAGCGTCAGTGACGCTGACGGTACAGGCCAGACT 1631

Qy   1833 TTTATTGGCTGGGATAGTGCAGCAACAGCAACAGCTGTTGGACGTGGTCAAGAGACAACA 1892
            |||||| |||| ||||||||| || ||| |   |  || ||   |   |  ||  |||||
Db   1632 ATTATTGTCTGGTATAGTGCAACAGCAGAACAATTTGCTGAGGGCCATTGAGGCGCAACA 1691

Qy   1893 AGAATTGTTGCGACTGACCGTCTGGGGAACAAAGAACCTCCAGACTAGGGTCACTGCCAT 1952
            |  |||||| ||| || |||||||| |  ||| | |||||| | || |||   || |
Db   1692 GCATATGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAGTCCTGGCTGT 1751

Qy   1953 CGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATGCTTGGGGATGTGCGTTTAGACAAGT 2012
            || |  ||| |||||||  ||   ||  |||     ||||||  ||  |    | ||   |
Db   1752 GGAAAGATACCTAAAGGATCAACAGCTCCTGGGGTTTTGGGGTTGCTCTGGAAAACTCAT 1811

Qy   2013 CTGCCACACTACTGTACCATGGCCA-----------AATGCAAGTCTAACACCAAAGTG 2060
            |||  |||||||||  || |||                 |||  |  |||       |  ||
Db   1812 TTGCACCACTACTGTGCCTTGGAATGCTAGTTGGAGTAATAAATCTCTGGATGATATTTG 1871

Qy   2061 GAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGGAAGAAAATATAAC 2120
            ||| ||   ||| |||    |||||||  || |  |||||    |    |       |||
Db   1872 GAATAACATGACCTGGATGCAGTGGGAAAGAGAAATTGACAATTACACAAGCTTAATATA 1931

Qy   2121 AGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATGAATTACAAAAGTT 2180
            |  | |||||  |  | ||||  |||||||| |||||     | ||||||    | ||
Db   1932 CTCATTACTAGAAAAATCGCAAACCCAACAAGAAAAGAATGAACAAGAATTATTGGAATT 1991

Qy   2181 GAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGATAAAGTATATACA 2240
            | |||  ||||     ||  | ||||||||||||:|  |   |||| |   |||||| |
Db   1992 GGATAAATGGGCAAGTTTGTGGAATTGGTTTGACATAACAAATTGGCTGTGGTATATAAA 2051

Qy   2241 ATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGATCTATATAGTACA 2300
            |       |  ||||| | |||  |  ||  ||||||||||  |   | || |
Db   2052 AATATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGCTGTACTTTC 2111

Qy   2301 AATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGT 2338
            ||  |   ||   | |||||||| ||    |||.|||
Db   2112 TATAGTGAATAGAGTTAGGCAGGGATACTCACCATTGT 2149


Search completed: November 16, 2002, 08:01:02
Job time : 257.22 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 16, 2002, 07:33:33 ; Search time 67.1566 Seconds
(without alignments)
14531.422 Million cell updates/sec

Title: M33262_COPY_6695_9335
Perfect score: 2641
Sequence: 1 TCTAGAAGCATGCTATAACA.....ACTGACCTACCTACAATATG 2641

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 328367 seqs, 184756068 residues

Total number of hits satisfying chosen parameters: 656734

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2434.6	92.2	6978	10	US-09-872-733-15	Sequence 15, Appl
2	2429.6	92.0	2640	10	US-09-872-733-18	Sequence 18, Appl
3	1745.4	66.1	2658	9	US-10-077-294-2	Sequence 2, Appli
4	242	9.2	2571	9	US-10-077-294-3	Sequence 3, Appli
5	242	9.2	2571	10	US-09-999-183-3	Sequence 3, Appli
6	232.6	8.8	9719	10	US-09-737-190A-1	Sequence 1, Appli
7	226.2	8.6	2583	9	US-10-003-035-24	Sequence 24, Appl
8	226.2	8.6	3157	9	US-10-003-035-14	Sequence 14, Appl
9	226.2	8.6	8933	10	US-09-943-286-3	Sequence 3, Appli
10	226.2	8.6	8933	10	US-09-943-286-4	Sequence 4, Appli
11	226.2	8.6	8933	10	US-09-943-286-9	Sequence 9, Appli
12	222.8	8.4	9544	10	US-09-798-675-4	Sequence 4, Appli
13	220.2	8.3	2950	9	US-10-003-035-22	Sequence 22, Appl
14	220.2	8.3	3417	9	US-10-003-035-21	Sequence 21, Appl
15	218.2	8.3	2280	9	US-10-003-035-16	Sequence 16, Appl
16	218.2	8.3	2747	9	US-10-003-035-23	Sequence 23, Appl
17	215	8.1	1568	10	US-09-756-551A-7	Sequence 7, Appli
18	212.6	8.0	3839	9	US-10-003-035-52	Sequence 52, Appl
19	212.2	8.0	1860	10	US-09-891-609-3	Sequence 3, Appli
20	212.2	8.0	1941	10	US-09-891-609-1	Sequence 1, Appli
21	208.6	7.9	2941	9	US-10-003-035-19	Sequence 19, Appl
22	208.2	7.9	4040	9	US-10-003-035-54	Sequence 54, Appl
23	206.6	7.8	2746	9	US-10-003-035-20	Sequence 20, Appl
24	193.2	7.3	2532	10	US-09-991-258-18	Sequence 18, Appl
25	193.2	7.3	13584	10	US-09-991-258-17	Sequence 17, Appl
26	190.2	7.2	9918	10	US-09-798-675-5	Sequence 5, Appli
27	189.6	7.2	2562	9	US-10-003-035-32	Sequence 32, Appl
28	176	6.7	9793	9	US-09-886-156-56	Sequence 56, Appl
29	151.4	5.7	2571	10	US-09-999-183-4	Sequence 4, Appli
30	135.6	5.1	2310	10	US-09-476-242-3	Sequence 3, Appli
31	135.6	5.1	2316	10	US-09-476-242-4	Sequence 4, Appli
32	135.6	5.1	2316	10	US-09-476-242-8	Sequence 8, Appli
33	135.6	5.1	2322	10	US-09-476-242-5	Sequence 5, Appli
34	135.6	5.1	2328	10	US-09-476-242-6	Sequence 6, Appli
35	135.6	5.1	2334	10	US-09-476-242-7	Sequence 7, Appli
36	135.6	5.1	2358	10	US-09-476-242-25	Sequence 25, Appl
37	135	5.1	1057	9	US-09-886-156-44	Sequence 44, Appl
38	134.2	5.1	2352	10	US-09-476-242-26	Sequence 26, Appl
c 39	133.4	5.1	1057	9	US-09-886-156-45	Sequence 45, Appl
40	132.6	5.0	2298	10	US-09-476-242-22	Sequence 22, Appl
41	132.6	5.0	2298	10	US-09-476-242-23	Sequence 23, Appl
42	132.6	5.0	2298	10	US-09-476-242-24	Sequence 24, Appl
43	132.6	5.0	2310	10	US-09-476-242-21	Sequence 21, Appl
44	132.6	5.0	2322	10	US-09-476-242-18	Sequence 18, Appl
45	132.6	5.0	2322	10	US-09-476-242-19	Sequence 19, Appl

ALIGNMENTS

RESULT 1
US-09-872-733-15
; Sequence 15, Application US/09872733
; Patent No. US20010036655A1
; GENERAL INFORMATION:
; APPLICANT: The Government of the United States of America, as
; TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, SIV GAG AND
; TITLE OF INVENTION: SIV ENV GENES
; FILE REFERENCE: 2026-4287US1 HIV GAG/POL,SIV GAG & ENV
; CURRENT APPLICATION NUMBER: US/09/872,733
; CURRENT FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: PCT/US00/34985
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/173,036
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 6978
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: DNA sequence
; OTHER INFORMATION: of construct CMVkan/R-R-SIVgp160 CTE
US-09-872-733-15

Query Match 92.2%; Score 2434.6; DB 10; Length 6978;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 2452; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

```
Qy    161 TAAGTATGGGATGTCTTGGGAATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATG 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    759 TAAGTATGGGATGTCTTGGGAATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATG 818

Qy    221 GGATCTATTGTACTCTATATGTCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGA 280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    819 GGATCTATTGTACTCTATATGTCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGA 878

Qy    281 CAATTCCCCTCTTTTGTGCAACCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTAC 340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    879 CAATTCCCCTCTTTTGTGCAACCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTAC 938
```

```
Qy    341 CAGATAATGGTGATTATTCAGAAGTGGCCCTTAATGTTACAGAAAGCTTTGATGCCTGGA 400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    939 CAGATAATGGTGATTATTCAGAAGTGGCCCTTAATGTTACAGAAAGCTTTGATGCCTGGA 998

Qy    401 ATAATACAGTCACAGAACAGGCAATAGAGGATGTATGGCAACTCTTTGAGACCTCAATAA 460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    999 ATAATACAGTCACAGAACAGGCAATAGAGGATGTATGGCAACTCTTTGAGACCTCAATAA 1058

Qy    461 AGCCTTGTGTAAAATTATCCCCATTATGCATTACTATGAGATGCAATAAAAGTGAGACAG 520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1059 AGCCTTGTGTAAAATTATCCCCATTATGCATTACTATGAGATGCAATAAAAGTGAGACAG 1118

Qy    521 ATAGATGGGGATTGACAAAATCAATAACAACAACAGCATCAACAACATCAACGACAGCAT 580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1119 ATAGATGGGGATTGACAAAATCAATAACAACAACAGCATCAACAACATCAACGACAGCAT 1178

Qy    581 CAGCAAAAGTAGACATGGTCAATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACAG 640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1179 CAGCAAAAGTAGACATGGTCAATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACAG 1238

Qy    641 GCTTGGAACAAGAGCAAATGATAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACA 700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1239 GCTTGGAACAAGAGCAAATGATAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACA 1298

Qy    701 AGAAAAAAGAGTACAATGAAACTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATA 760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1299 AGAAAAAAGAGTACAATGAAACTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATA 1358

Qy    761 ACACTGGTAATGAAAGTAGATGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGT 820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1359 ACACTGGTAATGAAAGTAGATGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGT 1418

Qy    821 CTTGTGACAAACATTATTGGGATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATG 880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1419 CTTGTGACAAACATTATTGGGATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATG 1478

Qy    881 CTTTGCTTAGATGTAATGACACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGG 940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1479 CTTTGCTTAGATGTAATGACACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGG 1538

Qy    941 TGGTCTCTTCATGCACAAGGATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATG 1000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1539 TGGTCTCTTCATGCACAAGGATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATG 1598

Qy   1001 GAACTAGAGCAGAAAATAGAACTTATATTTACTGGCATGGTAGGGATAATAGGACTATAA 1060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1599 GAACTAGAGCAGAAAATAGAACTTATATTTACTGGCATGGTAGGGATAATAGGACTATAA 1658

Qy   1061 TTAGTTTAAATAAGTATTATAATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAG 1120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1659 TTAGTTTAAATAAGTATTATAATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAG 1718

Qy   1121 TTTTACCAGTCACCATTATGTCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGGC 1180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1719 TTTTACCAGTCACCATTATGTCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGGC 1778

Qy   1181 CAAAGCAGGCATGGTGTTGGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGC 1240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1779 CAAAGCAGGCATGGTGTTGGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGC 1838

Qy   1241 AGACCATTGTCAAACATCCCAGGTATACTGGAACTAACAATACTGATAAAATCAATTTGA 1300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1839 AGACCATTGTCAAACATCCCAGGTATACTGGAACTAACAATACTGATAAAATCAATTTGA 1898

Qy   1301 CGGCTCCTGGAGGAGGAGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGT 1360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1899 CGGCTCCTGGAGGAGGAGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGT 1958

Qy   1361 TCCTCTACTGTAAAATGAATTGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACC 1420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1959 TCCTCTACTGTAAAATGAATTGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACC 2018

Qy   1421 AGAAGCCAAAGGAACAGCATAAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCA 1480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2019 AGAAGCCAAAGGAACAGCATAAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCA 2078

Qy   1481 ACACTTGGCATAAAGTAGGCAAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGT 1540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2079 ACACTTGGCATAAAGTAGGCAAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGT 2138

Qy   1541 GTAACTCCACAGTGACCAGTCTCATAGCAAACATAGATTGGATTGATGGAAACCAAACTA 1600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2139 GTAACTCCACAGTGACCAGTCTCATAGCAAACATAGATTGGATTGATGGAAACCAAACTA 2198

Qy   1601 ATATCACCATGAGTGCAGAGGTGGCAGAACTGTATCGATTGGAATTGGGAGATTATAAAT 1660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2199 ATATCACCATGAGTGCAGAGGTGGCAGAACTGTATCGATTGGAATTGGGAGATTATAAAT 2258

Qy   1661 TAGTAGAGATCACTCCAATTGGCTTGGCCCCCACAGATGTGAAGAGGTACACTACTGGTG 1720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2259 TAGTAGAGATCACTCCAATTGGCTTGGCCCCCACAGATGTGAAGAGGTACACTACTGGTG 2318

Qy   1721 GCACCTCAAGAAATAAAAGAGGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGG 1780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2319 GCACCTCAAGAAATAAAAGAGGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGG 2378

Qy   1781 CAGGTTCTGCAATGGGCGCGGCGTCGTTGACGCTGACCGCTCAGTCCCGAACTTTATTGG 1840
          |||||||||||||||| || ||    |||| || || || ||||||||||||||||||||
Db   2379 CAGGTTCTGCAATGGGAGCCGCCAGCCTGACCCTCACGGCACAGTCCCGAACTTTATTGG 2438

Qy   1841 CTGGGATAGTGCAGCAACAGCAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAATTGT 1900
          |||||||||| || || ||||||||||||||||||||||||||||||||||||||||||||
Db   2439 CTGGGATAGTCCAACAGCAGCAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAATTGT 2498

Qy   1901 TGCGACTGACCGTCTGGGGAACAAAGAACCTCCAGACTAGGGTCACTGCCATCGAGAAGT 1960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2499 TGCGACTGACCGTCTGGGGAACAAAGAACCTCCAGACTAGGGTCACTGCCATCGAGAAGT 2558

Qy   1961 ACTTAAAGGACCAGGCGCAGCTGAATGCTTGGGGATGTGCGTTTAGACAAGTCTGCCACA 2020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2559 ACTTAAAGGACCAGGCGCAGCTGAATGCTTGGGGATGTGCGTTTAGACAAGTCTGCCACA 2618

Qy   2021 CTACTGTACCATGGCCAAATGCAAGTCTAACACCAAAGTGGAACAATGAGACTTGGCAAG 2080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2619 CTACTGTACCATGGCCAAATGCAAGTCTAACACCAAAGTGGAACAATGAGACTTGGCAAG 2678

Qy   2081 AGTGGGAGCGAAAGGTTGACTTCTTGGAAGAAAATATAACAGCCCTCCTAGAGGAGGCAC 2140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2679 AGTGGGAGCGAAAGGTTGACTTCTTGGAAGAAAATATAACAGCCCTCCTAGAGGAGGCAC 2738

Qy   2141 AAATTCAACAAGAGAAGAACATGTATGAATTACAAAAGTTGAATAGCTGGGATGTGTTTG 2200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2739 AAATTCAACAAGAGAAGAACATGTATGAATTACAAAAGTTGAATAGCTGGGATGTGTTTG 2798

Qy   2201 GCAATTGGTTTGACCTTGCTTCTTGGATAAAGTATATACAATATGGAGTTTATATAGTTG 2260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2799 GCAATTGGTTTGACCTTGCTTCTTGGATAAAGTATATACAATATGGAGTTTATATAGTTG 2858

Qy   2261 TAGGAGTAATACTGTTAAGAATAGTGATCTATATAGTACAAATGCTAGCTAAGTTAAGGC 2320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2859 TAGGAGTAATACTGTTAAGAATAGTGATCTATATAGTACAAATGCTAGCTAAGTTAAGGC 2918

Qy   2321 AGGGGTATAGGCCAGTGTTCTCTTCCCCACCCTCTTATTTCCAGCAGACCCATATCCAAC 2380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2919 AGGGGTATAGGCCAGTGTTCTCTTCCCCACCCTCTTATTTCCAGCAGACCCATATCCAAC 2978

Qy   2381 AGGACCCGGCACTGCCAACCAGAGAAGGCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCA 2440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2979 AGGACCCGGCACTGCCAACCAGAGAAGGCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCA 3038

Qy   2441 ACAGCTCCTGGCCTTGGCAGATAGAATATATTCATTTCCTGATCCGCCAACTGATACGCC 2500
          |||||||||||||||||||||||||||||| || || || || || || || ||  | |
Db   3039 ACAGCTCCTGGCCTTGGCAGATAGAATATATCCACTTTCTTATTCGTCAGCTTATTAGAC 3098

Qy   2501 TCTTGACTTGGCTATTCAGCAACTGCAGAACCTTGCTATCGAGAGTATACCAGATCCTCC 2560
          |||||||||||||||||| ||||| || || ||||||||||||||||||||||||||||||
```

```
Db   3099 TCTTGACTTGGCTATTCAGTAACTGTAGGACTTTGCTATCGAGAGTATACCAGATCCTCC 3158

Qy   2561 AACCAATACTCCAGAGGCTCTCTGCGACCCTACAGAGGATTCGAGAAGTCCTCAGGACTG 2620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3159 AACCAATACTCCAGAGGCTCTCTGCGACCCTACAGAGGATTCGAGAAGTCCTCAGGACTG 3218

Qy   2621 AACTGACCTACCTACAATATG 2641
          |||||||||||||||||||||
Db   3219 AACTGACCTACCTACAATATG 3239


RESULT 2
US-09-872-733-18
; Sequence 18, Application US/09872733
; Patent No. US20010036655A1
; GENERAL INFORMATION:
;  APPLICANT: The Government of the United States of America, as
;  TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, SIV GAG AND
;  TITLE OF INVENTION:  SIV ENV GENES
;  FILE REFERENCE: 2026-4287US1 HIV GAG/POL,SIV GAG & ENV
;  CURRENT APPLICATION NUMBER: US/09/872,733
;  CURRENT FILING DATE:  2001-06-01
;  PRIOR APPLICATION NUMBER: PCT/US00/34985
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/173,036
;  PRIOR FILING DATE: 1999-12-23
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
;   LENGTH: 2640
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  DNA sequence
;   OTHER INFORMATION: of mutated SIV gene in construct
;   OTHER INFORMATION: CMVkan/R-R-SIVgp160 CTE
US-09-872-733-18

  Query Match             92.0%;  Score 2429.6;  DB 10;  Length 2640;
  Best Local Similarity   98.8%;  Pred. No. 0;
  Matches 2447;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy    166 ATGGGATGTCTTGGGAATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATC 225
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGGGATGTCTTGGGAATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATC 60

Qy    226 TATTGTACTCTATATGTCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATT 285
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TATTGTACTCTATATGTCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATT 120

Qy    286 CCCCTCTTTTGTGCAACCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGAT 345
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CCCCTCTTTTGTGCAACCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGAT 180

Qy    346 AATGGTGATTATTCAGAAGTGGCCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAAT 405
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AATGGTGATTATTCAGAAGTGGCCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAAT 240

Qy    406 ACAGTCACAGAACAGGCAATAGAGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCT 465
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ACAGTCACAGAACAGGCAATAGAGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCT 300

Qy    466 TGTGTAAAATTATCCCCATTATGCATTACTATGAGATGCAATAAAAGTGAGACAGATAGA 525
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TGTGTAAAATTATCCCCATTATGCATTACTATGAGATGCAATAAAAGTGAGACAGATAGA 360

Qy    526 TGGGGATTGACAAAATCAATAACAACAACAGCATCAACAACATCAACGACAGCATCAGCA 585
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TGGGGATTGACAAAATCAATAACAACAACAGCATCAACAACATCAACGACAGCATCAGCA 420

Qy    586 AAAGTAGACATGGTCAATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTG 645
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AAAGTAGACATGGTCAATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTG 480

Qy    646 GAACAAGAGCAAATGATAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAA 705
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GAACAAGAGCAAATGATAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAA 540

Qy    706 AAAGAGTACAATGAAACTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACT 765
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 AAAGAGTACAATGAAACTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACT 600

Qy    766 GGTAATGAAAGTAGATGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGT 825
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GGTAATGAAAGTAGATGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGT 660

Qy    826 GACAAACATTATTGGGATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTG 885
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GACAAACATTATTGGGATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTG 720

Qy    886 CTTAGATGTAATGACACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTC 945
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 CTTAGATGTAATGACACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTC 780

Qy    946 TCTTCATGCACAAGGATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACT 1005
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 TCTTCATGCACAAGGATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACT 840

Qy   1006 AGAGCAGAAAATAGAACTTATATTTACTGGCATGGTAGGGATAATAGGACTATAATTAGT 1065
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 AGAGCAGAAAATAGAACTTATATTTACTGGCATGGTAGGGATAATAGGACTATAATTAGT 900

Qy   1066 TTAAATAAGTATTATAATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTA 1125
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 TTAAATAAGTATTATAATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTA 960

Qy   1126 CCAGTCACCATTATGTCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGGCCAAAG 1185
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 CCAGTCACCATTATGTCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGGCCAAAG 1020

Qy   1186 CAGGCATGGTGTTGGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACC 1245
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 CAGGCATGGTGTTGGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACC 1080

Qy   1246 ATTGTCAAACATCCCAGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGCT 1305
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 ATTGTCAAACATCCCAGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGCT 1140

Qy   1306 CCTGGAGGAGGAGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTC 1365
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CCTGGAGGAGGAGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTC 1200

Qy   1366 TACTGTAAAATGAATTGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAG 1425
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 TACTGTAAAATGAATTGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAG 1260

Qy   1426 CCAAAGGAACAGCATAAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCAACACT 1485
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 CCAAAGGAACAGCATAAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCAACACT 1320

Qy   1486 TGGCATAAAGTAGGCAAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAAC 1545
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 TGGCATAAAGTAGGCAAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAAC 1380

Qy   1546 TCCACAGTGACCAGTCTCATAGCAAACATAGATTGGATTGATGGAAACCAAACTAATATC 1605
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 TCCACAGTGACCAGTCTCATAGCAAACATAGATTGGATTGATGGAAACCAAACTAATATC 1440

Qy   1606 ACCATGAGTGCAGAGGTGGCAGAACTGTATCGATTGGAATTGGGAGATTATAAATTAGTA 1665
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 ACCATGAGTGCAGAGGTGGCAGAACTGTATCGATTGGAATTGGGAGATTATAAATTAGTA 1500

Qy   1666 GAGATCACTCCAATTGGCTTGGCCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACC 1725
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 GAGATCACTCCAATTGGCTTGGCCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACC 1560
```

```
Qy   1726 TCAAGAAATAAAAGAGGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGT 1785
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 TCAAGAAATAAAAGAGGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCTCGCAACGGCAGGT 1620

Qy   1786 TCTGCAATGGGCGCGGCGTCGTTGACGCTGACCGCTCAGTCCCGAACTTTATTGGCTGGG 1845
          |||||||||||  ||  ||     |||| ||  ||  ||  ||||||||||||||||||||||||
Db   1621 TCTGCAATGGGAGCCGCCAGCCTGACCCTCACGGCACAGTCCCGAACTTTATTGGCTGGG 1680

Qy   1846 ATAGTGCAGCAACAGCAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGA 1905
          |||||  ||  ||  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 ATAGTCCAACAGCAGCAACAGCTGTTGGACGTGGTCAAGAGACAACAAGAATTGTTGCGA 1740

Qy   1906 CTGACCGTCTGGGGAACAAAGAACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTA 1965
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 CTGACCGTCTGGGGAACAAAGAACCTCCAGACTAGGGTCACTGCCATCGAGAAGTACTTA 1800

Qy   1966 AAGGACCAGGCGCAGCTGAATGCTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACT 2025
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 AAGGACCAGGCGCAGCTGAATGCTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACT 1860

Qy   2026 GTACCATGGCCAAATGCAAGTCTAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGG 2085
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 GTACCATGGCCAAATGCAAGTCTAACACCAAAGTGGAACAATGAGACTTGGCAAGAGTGG 1920

Qy   2086 GAGCGAAAGGTTGACTTCTTGGAAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATT 2145
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 GAGCGAAAGGTTGACTTCTTGGAAGAAAATATAACAGCCCTCCTAGAGGAGGCACAAATT 1980

Qy   2146 CAACAAGAGAAGAACATGTATGAATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAAT 2205
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 CAACAAGAGAAGAACATGTATGAATTACAAAAGTTGAATAGCTGGGATGTGTTTGGCAAT 2040

Qy   2206 TGGTTTGACCTTGCTTCTTGGATAAAGTATATACAATATGGAGTTTATATAGTTGTAGGA 2265
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 TGGTTTGACCTTGCTTCTTGGATAAAGTATATACAATATGGAGTTTATATAGTTGTAGGA 2100

Qy   2266 GTAATACTGTTAAGAATAGTGATCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGG 2325
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 GTAATACTGTTAAGAATAGTGATCTATATAGTACAAATGCTAGCTAAGTTAAGGCAGGGG 2160

Qy   2326 TATAGGCCAGTGTTCTCTTCCCCACCCTCTTATTTCCAGCAGACCCATATCCAACAGGAC 2385
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 TATAGGCCAGTGTTCTCTTCCCCACCCTCTTATTTCCAGCAGACCCATATCCAACAGGAC 2220

Qy   2386 CCGGCACTGCCAACCAGAGAAGGCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGC 2445
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 CCGGCACTGCCAACCAGAGAAGGCAAAGAAAGAGACGGTGGAGAAGGCGGTGGCAACAGC 2280

Qy   2446 TCCTGGCCTTGGCAGATAGAATATATTCATTTCCTGATCCGCCAACTGATACGCCTCTTG 2505
          |||||||||||||||||||||||||  ||  ||  ||  ||  ||  ||  ||  |  ||||||
Db   2281 TCCTGGCCTTGGCAGATAGAATATATCCACTTTCTTATTCGTCAGCTTATTAGACTCTTG 2340

Qy   2506 ACTTGGCTATTCAGCAACTGCAGAACCTTGCTATCGAGAGTATACCAGATCCTCCAACCA 2565
          |||||||||||||||  |||||  ||  ||  ||||||||||||||||||||||||||||||||||
Db   2341 ACTTGGCTATTCAGTAACTGTAGGACTTTGCTATCGAGAGTATACCAGATCCTCCAACCA 2400

Qy   2566 ATACTCCAGAGGCTCTCTGCGACCCTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTG 2625
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 ATACTCCAGAGGCTCTCTGCGACCCTACAGAGGATTCGAGAAGTCCTCAGGACTGAACTG 2460

Qy   2626 ACCTACCTACAATATG 2641
          ||||||||||||||||
Db   2461 ACCTACCTACAATATG 2476


RESULT 3
US-10-077-294-2
; Sequence 2, Application US/10077294
; Patent No. US20020159979A1
;   GENERAL INFORMATION:
;        APPLICANT: Johnson, Philip R.
;        TITLE OF INVENTION: Adeno-Associated Virus Materials and
;                            Methods
;        NUMBER OF SEQUENCES: 3
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
;             STREET: 6300 Sears Tower, 233 S. Wacker Drive
;             CITY: Chicago
;             STATE: Illinois
;             COUNTRY: USA
;             ZIP: 60606
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.25
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/077,294
;             FILING DATE: 15-Feb-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 09/691,604
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: No. US20020159979A1and, Greta E.
;             REGISTRATION NUMBER: 35,302
;             REFERENCE/DOCKET NUMBER: 31975
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (312) 474-6300
;             TELEFAX: (312) 474-0448
;             TELEX: 25-3856
;   INFORMATION FOR SEQ ID NO: 2:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 2658 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-077-294-2

  Query Match             66.1%;  Score 1745.4;  DB 9;  Length 2658;
  Best Local Similarity   82.2%;  Pred. No. 0;
  Matches 2052;  Conservative    0;  Mismatches  421;  Indels   24;  Gaps    3;

Qy    166 ATGGGATGTCTTGGGAATCAGCTGCTTATCGCCATCTTGCTTTTAAGTGTCTATGGGATC 225
          |||||||||||||||||||||||||||||||  |||||||  |||||||  |  | |||
Db      1 ATGGGATGTCTTGGGAATCAGCTGCTTATCGCGCTCTTGCTAGTAAGTGTTTTAGAGATT 60

Qy    226 TATTGTACTCTATATGTCACAGTCTTTTATGGTGTACCAGCTTGGAGGAATGCGACAATT 285
          | ||||  ||  ||||||  |||||  ||  ||||||||||||||  ||||  ||||||||||||
Db     61 TGTTGTGTTCAATATGTAACAGTATTCTATGGTGTACCAGCATGGAAGAATGCGACAATT 120

Qy    286 CCCCTCTTTTGTGCAACCAAGAATAGGGATACTTGGGGAACAACTCAGTGCCTACCAGAT 345
          ||||||||  ||||||||||||||||||||  ||||||||||||||  ||  |||  |  ||||||
Db    121 CCCCTCTTCTGTGCAACCAAGAATAGGGACACTTGGGGAACAACACAATGCTTGCCAGAT 180

Qy    346 AATGGTGATTATTCAGAAGTGGCCCTTAATGTTACAGAAAGCTTTGATGCCTGGAATAAT 405
          ||||  ||||||  ||||||  ||||  |  |||||  |||||     ||||||||  |||  |||||
Db    181 AATGATGATTACTCAGAATTGGCAATCAATGTCACAGAGGCTTTTGATGCTTGGGATAAT 240

Qy    406 ACAGTCACAGAACAGGCAATAGAGGATGTATGGCAACTCTTTGAGACCTCAATAAAGCCT 465
          ||||||||||||||  ||||||||||||||  |||  |  ||||||||  ||  ||  ||  |||||
Db    241 ACAGTCACAGAACAAGCAATAGAGGATGTGTGGAACCTCTTTGAAACATCCATTAAGCCC 300

Qy    466 TGTGTAAAATTATCCCCATTATGCATTACTATGAGATGCAATAAAAGTGAGACAGATAGA 525
          |||||||||  |  |||||  ||||  ||  |  ||||||||  |||||||  ||||||||||||
Db    301 TGTGTAAAACTCACCCCACTATGTATAGCAATGAGATGTAATAAAACTGAGACAGATAGG 360

Qy    526 TGGGGATTGACAAAATCA------ATAACAACAACAGCATCAACAACATCAACGACAGCA 579
          |||||  ||||||  |          | ||||||||||||  |||||||  ||  |  |||  ||
Db    361 TGGGGTTTGACAGGAAACGCAGGGACAACAACAACAGCAATAACAACAACAGCAACACCA 420

Qy    580 TCAGCAAAAGTAGACATGGTCAATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACA 639
            |  |  ||  |  |  |  |  |||||  |  ||  |  |  |  |||    |  |||  |||  ||
Db    421 AGTGTAGCAGAAAATGTTATAAATGAAAGTAATCCGGGCATAAAAAATAATAGTTGTGCA 480
```

```
Qy    640 GGCTTGGAACAAGAGCAAATGATAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGAC 699
Db    481 GGCTTGGAACAGGAGCCCATGATAGGTTGTAAATTTAACATGACAGGGTTAAATAGGGAC 540

Qy    700 AAGAAAAAAGAGTACAATGAAACTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAAT 759
Db    541 AAAAAGAAAGAATATAATGAAACATGGTATTCAAGAGATTTAATCTGTGAGCA---GTCA 597

Qy    760 AACACTGGTAATGAAAGTAGATGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAG 819
Db    598 GCGAATGAAAGTGAGAGTAAATGTTACATGCATCATTGTAACACCAGTGTTATTCAAGAA 657

Qy    820 TCTTGTGACAAACATTATTGGGATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTAT 879
Db    658 TCCTGTGACAAGCATTATTGGGATGCTATTAGATTTAGATACTGTGCACCGCCAGGTTAT 717

Qy    880 GCTTTGCTTAGATGTAATGACACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTG 939
Db    718 GCTTTGCTTAGGTGTAATGATTCAAATTATTTAGGCTTTGCTCCTAACTGTTCTAAGGTA 777

Qy    940 GTGGTCTCTTCATGCACAAGGATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAAT 999
Db    778 GTGGTTTCTTCATGCACAAGAATGATGGAGACGCAAACCTCTACTTGGTTTGGCTTCAAT 837

Qy   1000 GGAACTAGAGCAGAAAATAGAACTTATATTTACTGGCATGGTAGGGATAATAGGACTATA 1059
Db    838 GGTACTAGGGCAGAAAATAGAACATACATTTATTGGCATGGCAAAAGTAATAGAACCATA 897

Qy   1060 ATTAGTTTAAATAAGTATTATAATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACA 1119
Db    898 ATTAGCTTGAATAAGTATTATAATCTAACAATGAGATGTAGAAGACCAGAAAATAAGACA 957

Qy   1120 GTTTTACCAGTCACCATTATGTCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGG 1179
Db    958 GTTTTACCAGTCACCATTATGTCAGGGTTGGTCTTCCATTCGCAGCCCATAAATGAGAGA 1017

Qy   1180 CCAAAGCAGGCATGGTGTTGGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAG 1239
Db   1018 CCAAAACAGGCCTGGTGCTGGTTTGAAGGAAGCTGGAAAAAGGCCATCCAGGAAGTGAAG 1077

Qy   1240 CAGACCATTGTCAAACATCCCAGGTATACTGGAACTAACAATACTGATAAAATCAATTTG 1299
Db   1078 GAAACCTTGGTCAAACATCCCAGGTATACGGGAACTAATGATACTAGGAAAATTAATCTA 1137

Qy   1300 ACGGCTCCTGGAGGAGGAGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAG 1359
Db   1138 ACAGCTCCAGCAGGAGGAGATCCAGAAGTCACTTTTATGTGGACAAATTGTCGAGGAGAA 1197

Qy   1360 TTCCTCTACTGTAAAATGAATTGGTTTCTAAATTGGGTAGAAGATAGGAATACA------ 1413
Db   1198 TTCTTATATTGCAAAATGAATTGGTTTCTTAATTGGGTAGAGGACAGAGACCAAAAGGGT 1257

Qy   1414 ---------GCTAACCAGAAGCCAAAGGAACAGCATAAAAGGAATTACGTGCCATGTCAT 1464
Db   1258 GGCAGATGGAAACAACAAAATAGGAAAGAGCAACAGAAGAAAAATTATGTGCCATGTCAT 1317

Qy   1465 ATTAGACAAATAATCAACACTTGGCATAAAGTAGGCAAAAATGTTTATTTGCCTCCAAGA 1524
Db   1318 ATTAGACAAATAATCAACACGTGGCACAAAGTAGGCAAAAATGTATATTTGCCTCCTAGG 1377

Qy   1525 GAGGGAGACCTCACGTGTAACTCCACAGTGACCAGTCTCATAGCAAACATAGATTGGATT 1584
Db   1378 GAAGGAGACCTGACATGCAATTCCACTGTAACTAGTCTCATAGCAGAGATAGATTGGATC 1437

Qy   1585 GATGGAAACCAAACTAATATCACCATGAGTGCAGAGGTGGCAGAACTGTATCGATTGGAA 1644
Db   1438 AATAGCAATGAGACCAATATCACCATGAGTGCAGAGGTGGCAGAACTGTATCGATTGGAG 1497

Qy   1645 TTGGGAGATTATAAATTAGTAGAGATCACTCCAATTGGCTTGGCCCCCACAGATGTGAAG 1704
Db   1498 TTGGGAGATTACAAATTAATAGAGATTACTCCAATTGGCTTGGCCCCCACAAGTGTAAGA 1557

Qy   1705 AGGTACACTACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCTTTGTGCTAGGGTTCTTG 1764
Db   1558 AGGTACACCACAACTGGTGCCTCAAGAAATAAGAGAGGGGTCTTTGTGCTAGGGTTCTTG 1617

Qy   1765 GGTTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGTTGACGCTGACCGCTCAG 1824
Db   1618 GGTTTTCTCGCGACAGCAGGTTCTGCAATGGGCGCGGCGTCCGTGACGCTGTCGGCTCAG 1677

Qy   1825 TCCCGAACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGCTGTTGGACGTGGTCAAG 1884
Db   1678 TCCCGGACTTTGTTGGCTGGGATAGTGCAGCAACAGCAACAGCTGTTGGATGTGGTCAAG 1737

Qy   1885 AGACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACAAAGAACCTCCAGACTAGGGTC 1944
Db   1738 AGACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACTAAGAACCTCCAGACTAGAGTC 1797

Qy   1945 ACTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATGCTTGGGGATGTGCGTTT 2004
Db   1798 ACTGCTATCGAGAAGTACCTGAAGGATCAGGCGCAGCTAAATTCATGGGGATGTGCTTTT 1857

Qy   2005 AGACAAGTCTGCCACACTACTGTACCATGGCCAAATGCAAGTCTAACACCAAAGTGGAAC 2064
Db   1858 AGGCAAGTCTGTCACACTACTGTACCATGGCCAAATGAAACATTGGTGCCTAATTGGAAC 1917

Qy   2065 AATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGGAAGAAAATATAACAGCC 2124
Db   1918 AATATGACTTGGCAAGAGTGGGAAAGACAGGTTGACTTCCTAGAGGCAAATATAACTCAA 1977

Qy   2125 CTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATGAATTACAAAAGTTGAAT 2184
Db   1978 TTATTAGAAGAAGCACAAATTCAGCAAGAAAAGAATATGTATGAATTGCAAAAATTAAAT 2037

Qy   2185 AGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGATAAAGTATATACAATAT 2244
Db   2038 AGCTGGGATATCTTTGGCAATTGGTTTGACCTTACTTCTTGGATAAGATATATACAATAT 2097

Qy   2245 GGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGATCTATATAGTACAAATG 2304
Db   2098 GGTGTACTTATAGTTCTAGGAGTAATAGGGTTAAGAATAGTAATATATGTAGTGCAAATG 2157

Qy   2305 CTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTCTCTTCCCCCACCCTCTTATTTCCAG 2364
Db   2158 TTAGCTAGGTTAAGACAGGGTTATAGGCCAGTGTTCTCTTCCCCTCCCGCTTATGTTCAG 2217

Qy   2365 CAGACCCATATCCAACAGGACCCGGCACTGCCAACCAGAGAAGGCAAAGAAAGAGACGGT 2424
Db   2218 CAGATCCCTATCCACAAGGGCCAGGAACCGCCAACCAAAGAAGGAGAAGAAGGAGACGGT 2277

Qy   2425 GGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAATATATTCATTTCCTGATC 2484
Db   2278 GGAGACAGAGGTGGCAGCAGATCTTGGCCTTGGCAGATAGAATATATTCATTTCCTGATC 2337

Qy   2485 CGCCAACTGATACGCCTCTTGACTTGGCTATTCAGCAACTGCAGAACCTTGCTATCGAGA 2544
Db   2338 CGCCAGTTGATACGCCTCTTGACTTGGCTATTCAGCAGCTGCAGGGATTGGCTATTGAGG 2397

Qy   2545 GTATACCAGATCCTCCAACCAATACTCCAGAGGCTCTCTGCGACCCTACAGAGGATTCGA 2604
Db   2398 AGCTACCAGATCCTCCAACCAGTGCTCCAGAGCCTCTCAACGACGTTGCAAAGAGTCCGT 2457

Qy   2605 GAAGTCCTCAGGACTGAACTGACCTACCTACAATATG 2641
Db   2458 GAAGTCATCAGAATTGAAATAGCCTACCTACAATATG 2494
```

RESULT 4
US-10-077-294-3
; Sequence 3, Application US/10077294
; Patent No. US20020159979A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Philip R.
; TITLE OF INVENTION: Adeno-Associated Virus Materials and
; Methods

```
;       NUMBER OF SEQUENCES: 3
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
;            STREET: 6300 Sears Tower, 233 S. Wacker Drive
;            CITY: Chicago
;            STATE: Illinois
;            COUNTRY: USA
;            ZIP: 60606
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible.
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.25
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/077,294
;            FILING DATE: 15-Feb-2002
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 09/691,604
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: No. US20020159979Aland, Greta E.
;            REGISTRATION NUMBER: 35,302
;            REFERENCE/DOCKET NUMBER: 31975
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (312) 474-6300
;            TELEFAX: (312) 474-0448
;            TELEX: 25-3856
;   INFORMATION FOR SEQ ID NO: 3:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 2571 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: DNA (genomic)
;       SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-10-077-294-3

  Query Match             9.2%;  Score 242;  DB 9;  Length 2571;
  Best Local Similarity   50.2%;  Pred. No. 1.8e-52;
  Matches  862;  Conservative    0;  Mismatches  790;  Indels   66;  Gaps    8;

Qy     651 AGAGCAAATGATAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGA 710
           ||   |||||| || |||    ||||| || ||       | |||||| ||||   |  |
Db     468 AGGAGAAATGAAAAACTGCTCTTTCAATATCACCACAAGCATAAGAGATAAGATGCAGAA 527

Qy     711 GTACAATGAAACTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAA 770
             |  ||| |  |   ||       ||| | ||||    | |     |||  ||
Db     528 AGAATATGCACTTCTTTATAAACTTGATATAGTATCAATAGATAATGATAGTAC------ 581

Qy     771 TGAAAGTAGATGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAA 830
             || | |    |||     ||||| || || || ||   | |  |||||   ||
Db     582 ------CAGCTATAGGTTGATAAGTTGTAATACCTCAGTCATTACACAAGCTTGTCCAAA 635

Qy     831 ACATTATTGGGATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAG 890
             |  |  ||  | |||    |    |||||||| ||  | |||| |||  | || |
Db     636 GATATCCTTTGAGCCAATTCCCATACACTATTGTGCCCCGGCTGGTTTTGCGATTCTAAA 695

Qy     891 ATGTAATGACACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTC 950
           ||||||  || | ||| |     ||        |   ||||   || ||
Db     696 ATGTAACGATAAAAAGTTCAGTGGAAAAGGATC---ATGTAAAAATGTCAGCACAGTACA 752

Qy     951 ATGCACAAGGATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGC 1010
           |||  |||      ||   | ||      || |||    |    || ||||| | |  |||
Db     753 ATGTACACATGGAATTAGGCCAGTAGTATCAACTCAACTGCTGTTAAATGGCAGTCTAGC 812

Qy    1011 AGAAAATAGAACTTATATTTACT---GGCATGGTAGGGATAATAGGACTATAATTAGTTT 1067
           ||||  |          |||   |    | ||   |  ||||||   |   |  || |   |
Db     813 AGAAGAAGAGGTAGTAATTAGATCTGAGAATTTCACTGATAATGCTAAAACCATCATAGT 872

Qy    1068 AAATAAGTATTATAATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACC 1127
           | ||  | || |   | ||  ||| || |||| ||||||      | || | |     |
Db     873 ACATCTGAATGAATCTGTACAAATTAATTGTACAAGACCCAACTACAATAAAAGAAAAAG 932

Qy    1128 AGTCACCATTATGTCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGGCCAAAGCA 1187
           |    ||     | ||    |  ||  |  || |   ||   || |||||    |  |
Db     933 GATACATATAGGACCAGGGAGAGCATTTTATACA-ACAAAAAATATAATAGGAACTATAA 991

Qy    1188 GGCATGGTGTTGGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCAT 1247
           | || |    | ||          |  | ||    | |  | |       ||   | |
Db     992 GACAAGCACATTGTAACATTAGTAGAGCAAAATGGAATGACACTTTAAGACAGATAGT-- 1049

Qy    1248 TGTCAAACATCCCAGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCC 1307
                    |  ||  |  |  | |       || | | | | |||    ||  |      |
Db    1050 ---------TAGCAAATTAAAAGAACAATTTAAGAATAAAACAATAGTCTTTAATCAATC 1100

Qy    1308 TGGAGGAGGAGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTA 1367
           ||||||  ||  ||  |||  ||     |      |   |||||  ||||  ||  ||  ||||
Db    1101 CTCAGGAGGGGACCCAGAAATTGTAATGCACAGTTTTAATTGTGGAGGGGAATTTTTCTA 1160

Qy    1368 CTGTAAAATGAATTGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCC 1427
           ||||||  |         |  | |||  |        || |  ||||    ||     |  ||  |
Db    1161 CTGTAATACATCACCACTGTTTAATAGTACTTGGAATGGTAATAATACTTGGAATAATAC 1220

Qy    1428 AAAGGAACAGCATAAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCAACACTTG 1487
           |  |      ||||  |  |       | |  |||   |  ||  |  ||||||  ||  ||||  ||
Db    1221 TACAGGGTCAAATAACAATATCACACTTCAATGCAAAATAAAACAAATTATAAACATGTG 1280

Qy    1488 GCATAAAGTAGGCAAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTC 1547
           |||  |||||||  |||     | |||   |||||  |   ||  |||  |   |  |||   ||
Db    1281 GCAGGAAGTAGGAAAAGCAATGTATGCCCCTCCCATTGAAGGACAAATTAGATGTTCATC 1340

Qy    1548 CACAGTGACCAGTCTCATAGCAAACATAGATTGGATTGATGGAAACCAAACTAATATCAC 1607
           |   | ||  | ||  ||  | ||||  |   | | |   |    | ||  ||   |||
Db    1341 AAATATTACAGGGCTACTATTAACAAGAGATGGTGGTAAGGACACGGACACGAACGACAC 1400

Qy    1608 CATGA---------------GTGCAGAGGTGGCAGAACTGTATCGATTGGAATTGGGAGA 1652
           |  ||                | | |||  ||   ||    |   ||   |||||     |
Db    1401 CGAGATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAA 1460

Qy    1653 TTATAAATTAGTAGAGATCACTCCAATTGGCTTGGCCCCCACAGATGTGAAGAGGTACAC 1712
           ||||||  |||||   ||     ||| | ||  | || |||||  | |  |||||
Db    1461 ATATAAAGTAGTAACAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAAGAG------ 1514

Qy    1713 TACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCT 1772
             | |         ||| | |||||||  |   | |        |||| | || ||  |
Db    1515 ---AAGAGTGGTGCAGAGAGAAAAAAGAGCAGCGATAGGAGCTCTGTTCCTTGGGTTCTT 1571

Qy    1773 CGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGTTGACGCTGACCGCTCAGTCCCGAAC 1832
           | | | |||||     | ||||||||  |||||  |||||||||| |  ||| || ||
Db    1572 AGGAGCAGCAGGAAGCACTATGGGCGCAGCGTCAGTGACGCTGACGGTACAGGCCAGACT 1631

Qy    1833 TTTATTGGCTGGGATAGTGCAGCAACAGCAACAGCTGTTGGACGTGGTCAAGAGACAACA 1892
           ||||||  |||| ||||||||  || ||| |  |  || ||   |   ||   |||||
Db    1632 ATTATTGTCTGGTATAGTGCAACAGCAGAACAATTTGCTGAGGGCCATTGAGGCGCAACA 1691

Qy    1893 AGAATTGTTGCGACTGACCGTCTGGGGAACAAAGAACCTCCAGACTAGGGTCACTGCCAT 1952
             |  |||||| ||| || |||||||| |  ||| | |||||| | || |||   ||  |
Db    1692 GCATATGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAGTCCTGGCTGT 1751

Qy    1953 CGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATGCTTGGGGATGTGCGTTTAGACAAGT 2012
           || |  ||| |||||||  ||   ||  |||     |||||| ||  |    | ||   |
Db    1752 GGAAAGATACCTAAAGGATCAACAGCTCCTGGGGTTTTGGGGTTGCTCTGGAAAACTCAT 1811

Qy    2013 CTGCCACACTACTGTACCATGGCCA------------AATGCAAGTCTAACACCAAAGTG 2060
           |||  |||||||||  || |||                |||  |  |||       |  ||
Db    1812 TTGCACCACTACTGTGCCTTGGAATGCTAGTTGGAGTAATAAATCTCTGGATGATATTTG 1871

Qy    2061 GAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGGAAGAAAATATAAC 2120
           ||| ||   ||| |||    |||||| || |  |||||    |    |      |||
Db    1872 GAATAACATGACCTGGATGCAGTGGGAAAGAGAAATTGACAATTACACAAGCTTAATATA 1931

Qy    2121 AGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATGAATTACAAAAGTT 2180
           |  | |||||  |  | ||||  |||||||| |||||     | ||||||     | ||
Db    1932 CTCATTACTAGAAAAATCGCAAACCCAACAAGAAAAGAATGAACAAGAATTATTGGAATT 1991
```

```
Qy   2181 GAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGATAAAGTATATACA 2240
          | |||  ||||     ||  | ||||||||||| |  |    ||||  |   ||||||| |
Db   1992 GGATAAATGGGCAAGTTTGTGGAATTGGTTTGACATAACAAATTGGCTGTGGTATATAAA 2051

Qy   2241 ATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGATCTATATAGTACA 2300
          |       |    ||||| | |||  |  ||   ||||||||||  |   | || |
Db   2052 AATATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGCTGTACTTTC 2111

Qy   2301 AATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGT 2338
           ||  |   ||   | |||||||| ||    ||| |||
Db   2112 TATAGTGAATAGAGTTAGGCAGGGATACTCACCATTGT 2149


RESULT 5
US-09-999-183-3
; Sequence 3, Application US/09999183
; Patent No. US20020147169A1
; GENERAL INFORMATION:
;  APPLICANT: MITROPHANOUS, et al
;  TITLE OF INVENTION: In Vivo Selection Method
;  FILE REFERENCE: 674523-2009
;  CURRENT APPLICATION NUMBER: US/09/999,183
;  CURRENT FILING DATE:  2001-11-29
;  PRIOR APPLICATION NUMBER: PCT/GB00/02136
;  PRIOR FILING DATE: 2000-06-02
;  PRIOR APPLICATION NUMBER: 9912965.2
;  PRIOR FILING DATE: 1999-06-03
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: SeqWin99
; SEQ ID NO 3
;   LENGTH: 2571
;   TYPE: DNA
;   ORGANISM: Human Immunodeficiency virus type I
US-09-999-183-3

  Query Match             9.2%;  Score 242;  DB 10;  Length 2571;
  Best Local Similarity   50.2%;  Pred. No. 1.8e-52;
  Matches  862;  Conservative    0;  Mismatches  790;  Indels   66;  Gaps    8;

Qy    651 AGAGCAAATGATAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGA 710
          ||   |||||| || |||    ||||| || ||    | ||||||| ||||   |  |
Db    468 AGGAGAAATGAAAAACTGCTCTTTCAATATCACCACAAGCATAAGAGATAAGATGCAGAA 527

Qy    711 GTACAATGAAACTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAA 770
            | ||| |  |   ||       ||| | ||||    | |     |||  ||
Db    528 AGAATATGCACTTCTTTATAAACTTGATATAGTATCAATAGATAATGATAGTAC------ 581

Qy    771 TGAAAGTAGATGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAA 830
              || | |    |||     ||||| || || || ||   | |  |||||   ||
Db    582 ------CAGCTATAGGTTGATAAGTTGTAATACCTCAGTCATTACACAAGCTTGTCCAAA 635

Qy    831 ACATTATTGGGATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAG 890
              |  ||  | |||     |    |||||||| ||  | |||| |||  | || |
Db    636 GATATCCTTTGAGCCAATTCCCATACACTATTGTGCCCCGGCTGGTTTTGCGATTCTAAA 695

Qy    891 ATGTAATGACACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTC 950
          ||||||  || | ||| |     ||       |   ||||   || ||
Db    696 ATGTAACGATAAAAAGTTCAGTGGAAAAGGATC---ATGTAAAAATGTCAGCACAGTACA 752

Qy    951 ATGCACAAGGATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGC 1010
          ||| |||      ||   | ||      || |||     |    || |||||  | |  |||
Db    753 ATGTACACATGGAATTAGGCCAGTAGTATCAACTCAACTGCTGTTAAATGGCAGTCTAGC 812

Qy   1011 AGAAAATAGAACTTATATTTACT---GGCATGGTAGGGATAATAGGACTATAATTAGTTT 1067
          ||||  |           |||   |    | ||   |  ||||||   |  |  || |   |
Db    813 AGAAGAAGAGGTAGTAATTAGATCTGAGAATTTCACTGATAATGCTAAAACCATCATAGT 872

Qy   1068 AAATAAGTATTATAATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACC 1127
          | ||  | || |   | ||  ||| || |||| ||||||      | || | |    |
Db    873 ACATCTGAATGAATCTGTACAAATTAATTGTACAAGACCCAACTACAATAAAAGAAAAAG 932

Qy   1128 AGTCACCATTATGTCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGGCCAAAGCA 1187
            |    ||     | ||    |  ||  |  || |   ||   || |||||    |  |

Db    933 GATACATATAGGACCAGGGAGAGCATTTTATACA-ACAAAAAATATAATAGGAACTATAA 991

Qy   1188 GGCATGGTGTTGGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCAT 1247
          | || |    | ||          |  | ||    |.|  | |     ||    | ||
Db    992 GACAAGCACATTGTAACATTAGTAGAGCAAAATGGAATGACACTTTAAGACAGATAGT-- 1049

Qy   1248 TGTCAAACATCCCAGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCC 1307
                    |  ||  |  |  | |       || | | | | |||     || |    |
Db   1050 ---------TAGCAAATTAAAAGAACAATTTAAGAATAAAACAATAGTCTTTAATCAATC 1100

Qy   1308 TGGAGGAGGAGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTA 1367
            ||||||  ||  || ||| ||     |      |  ||||  |||| || ||  ||||
Db   1101 CTCAGGAGGGGACCCAGAAATTGTAATGCACAGTTTTAATTGTGGAGGGGAATTTTTCTA 1160

Qy   1368 CTGTAAAATGAATTGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCC 1427
          ||||||  |          |  | ||| |         || | ||||   ||    | ||  |
Db   1161 CTGTAATACATCACCACTGTTTAATAGTACTTGGAATGGTAATAATACTTGGAATAATAC 1220

Qy   1428 AAAGGAACAGCATAAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCAACACTTG 1487
          |  |       |||| |  |       | | ||| | || | |||||| || ||||  ||
Db   1221 TACAGGGTCAAATAACAATATCACACTTCAATGCAAAATAAAACAAATTATAAACATGTG 1280

Qy   1488 GCATAAAGTAGGCAAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGTGTAACTC 1547
          |||  |||||||  |||    | |||   |||||  |  || ||| |  | |  |||  ||
Db   1281 GCAGGAAGTAGGAAAAGCAATGTATGCCCCTCCCATTGAAGGACAAATTAGATGTTCATC 1340

Qy   1548 CACAGTGACCAGTCTCATAGCAAACATAGATTGGATTGATGGAAACCAAACTAATATCAC 1607
            |  | ||  | ||  ||  ||  | |||| |   | | |   |   | || ||  | |||
Db   1341 AAATATTACAGGGCTACTATTAACAAGAGATGGTGGTAAGGACACGGACACGAACGACAC 1400

Qy   1608 CATGA---------------GTGCAGAGGTGGCAGAACTGTATCGATTGGAATTGGGAGA 1652
          |  ||               | | |||  ||   ||     ||   ||||  |   |
Db   1401 CGAGATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAA 1460

Qy   1653 TTATAAATTAGTAGAGATCACTCCAATTGGCTTGGCCCCCACAGATGTGAAGAGGTACAC 1712
          ||||||  |||||   ||    ||| | ||  | || |||||  | |  |||||
Db   1461 ATATAAAGTAGTAACAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAAGAG------ 1514

Qy   1713 TACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCTTTGTGCTAGGGTTCTTGGGTTTTCT 1772
            | |          ||| | |||||||  |   | |       ||||  | ||  ||  |
Db   1515 ---AAGAGTGGTGCAGAGAGAAAAAAGAGCAGCGATAGGAGCTCTGTTCCTTGGGTTCTT 1571

Qy   1773 CGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGTTGACGCTGACCGCTCAGTCCCGAAC 1832
          | | | ||||      | |||||||| ||||  ||||||||||  |  ||| ||||
Db   1572 AGGAGCAGCAGGAAGCACTATGGGCGCAGCGTCAGTGACGCTGACGGTACAGGCCAGACT 1631

Qy   1833 TTTATTGGCTGGGATAGTGCAGCAACAGCAACAGCTGTTGGACGTGGTCAAGAGACAACA 1892
          ||||||  |||| |||||||| || ||| |  |  || ||   |  ||  ||||||
Db   1632 ATTATTGTCTGGTATAGTGCAACAGCAGAACAATTTGCTGAGGGCCATTGAGGCGCAACA 1691

Qy   1893 AGAATTGTTGCGACTGACCGTCTGGGGAACAAAGAACCTCCAGACTAGGGTCACTGCCAT 1952
            |  |||||| ||| || |||||||| |  ||| | ||||||  | || |||   ||  |
Db   1692 GCATATGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAGTCCTGGCTGT 1751

Qy   1953 CGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATGCTTGGGGATGTGCGTTTAGACAAGT 2012
          || |  ||| |||||||  ||   ||  |||     ||||||  ||  |    | ||  |
Db   1752 GGAAAGATACCTAAAGGATCAACAGCTCCTGGGGTTTTGGGGTTGCTCTGGAAAACTCAT 1811

Qy   2013 CTGCCACACTACTGTACCATGGCCA----------AATGCAAGTCTAACACCAAAGTG 2060
          |||  |||||||||  || |||              |||  |  |||         ||
Db   1812 TTGCACCACTACTGTGCCTTGGAATGCTAGTTGGAGTAATAAATCTCTGGATGATATTTG 1871

Qy   2061 GAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGGAAGAAAATATAAC 2120
          ||| ||   ||| |||    |||||||  || |  |||||   |    |      |||
Db   1872 GAATAACATGACCTGGATGCAGTGGGAAAGAGAAATTGACAATTACACAAGCTTAATATA 1931

Qy   2121 AGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATGAATTACAAAAGTT 2180
          |  | |||||  |  | ||||  ||||||| |||||      | ||||||    | ||
Db   1932 CTCATTACTAGAAAAATCGCAAACCCAACAAGAAAAGAATGAACAAGAATTATTGGAATT 1991

Qy   2181 GAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGATAAAGTATATACA 2240
          | |||  ||||     ||  | ||||||||||| |  |    ||||  |   ||||||| |
Db   1992 GGATAAATGGGCAAGTTTGTGGAATTGGTTTGACATAACAAATTGGCTGTGGTATATAAA 2051
```

```
Qy   2241 ATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGATCTATATAGTACA 2300
Db   2052 AATATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGCTGTACTTTC 2111

Qy   2301 AATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGT 2338
Db   2112 TATAGTGAATAGAGTTAGGCAGGGATACTCACCATTGT 2149


RESULT 6
US-09-737-190A-1
; Sequence 1, Application US/09737190A
; Patent No. US20020102545A1
; GENERAL INFORMATION:
;  APPLICANT: Shibuya, Tetsuo
;  TITLE OF INVENTION: A Method for Changing a Target Array, a Method for Analyzing
;  TITLE OF INVENTION:  a Structure, and an Apparatus, a Storage Medium and a
;  TITLE OF INVENTION:  Transmission Medium Therefor
;  FILE REFERENCE: JP919990270US1 (14043)
;  CURRENT APPLICATION NUMBER: US/09/737,190A
;  CURRENT FILING DATE:  2000-12-14
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 9719
;   TYPE: DNA
;   ORGANISM: Human iImmunodeficiency virus type 1
US-09-737-190A-1

  Query Match             8.8%;  Score 232.6;  DB 10;  Length 9719;
  Best Local Similarity   48.9%;  Pred. No. 9.4e-50;
  Matches  909;  Conservative    0;  Mismatches  879;  Indels   69;  Gaps    8;

Qy    768 TAATGAAAGTAGATGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGA 827
Db   6782 TGATACTACCAGCTATAAGTTGACAAGTTGTAACACCTCAGTCATTACACAGGCCTGTCC 6841

Qy    828 CAAACATTATTGGGATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCT 887
Db   6842 AAAGGTATCCTTTGAGCCAATTCCCATACATTATTGTGCCCCGGCTGGTTTTGCGATTCT 6901

Qy    888 TAGATGTAATGACACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTC 947
Db   6902 AAAATGTAATAATAAGACGTTCAATGGAACAGGACC---ATGTACAAATGTCAGCACAGT 6958

Qy    948 TTCATGCACAAGGATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAG 1007
Db   6959 ACAATGTACACATGGAATTAGGCCAGTAGTATCAACTCAACTGCTGTTAAATGGCAGTCT 7018

Qy   1008 AGCAGAAAATAGAACTTATATTTACTGGC---ATGGTAGGGATAATAGGACTATAATTAG 1064
Db   7019 AGCAGAAGAAGAGGTAGTAATTAGATCTGTCAATTTCACGGACAATGCTAAAACCATAAT 7078

Qy   1065 TTTAAATAAGTATTATAATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTT 1124
Db   7079 AGTACAGCTGAACACATCTGTAGAAATTAATTGTACAAGACCCAACAACAATACAAGAAA 7138

Qy   1125 ACCAGTCACCATTATGTCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGGCCAAA 1184
Db   7139 AAGAATCCGTATCCAGAGAGGACCAGGGAGAGCATTTGTTACAATAGGAAAAATAGGAAA 7198

Qy   1185 GCAGGCATGGTGTTGGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGAC 1244
Db   7199 TATGAGACAAGCACATTGTAACATTAGTAGAGCAAAATGGAATAACACTTTAAAACAGAT 7258

Qy   1245 CATTGTCAAACATCCCAGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGC 1304
Db   7259 AGCTAGCAAATTAAGAGAACAATTTGGAAATAATAAAAC------AATAATCTTTAAGCA 7312

Qy   1305 TCCTGGAGGAGGAGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCT 1364
Db   7313 ATCCTCAGGAGGGGACCCAGAAATTGTAACGCACAGTTTTAATTGTGGAGGGGAATTTTT 7372

Qy   1365 CTACTGTAAAATGAATTGGTTTCTAAATTGGGTAGAAGATA---------GGAATACAGC 1415
Db   7373 CTACTGTAATTCAACACAACTGTTTAATAGTACTTGGTTTAATAGTACTTGGAGTACTGA 7432

Qy   1416 TAACCAGAAGCCAAAGGAACAGCATAAAAGGAATTACGTGCCATGTCATATTAGACAAAT 1475
Db   7433 AGGGTCAAATAACACTGAAGGAAGTGACACAATCACCCTCCCATGCAGAATAAAACAAAT 7492

Qy   1476 AATCAACACTTGGCATAAAGTAGGCAAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCT 1535
Db   7493 TATAAACATGTGGCAGAAAGTAGGAAAAGCAATGTATGCCCCTCCCATCAGTGGACAAAT 7552

Qy   1536 CACGTGTAACTCCACAGTGACCAGTCTCATAGCAAACATAGATTGGATTGATGGAAACCA 1595
Db   7553 TAGATGTTCATCAAATATTACAGGGCTGCTATTAACAAGAGATGGTGGTAATAGCAACAA 7612

Qy   1596 AACTAATATCACCATGAGTGCAGAGGTGGCAGAACTG---------TATCGATTGGAATT 1646
Db   7613 TGAGTCCGAGATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATT 7672

Qy   1647 GGGAGATTATAAATTAGTAGAGATCACTCCAATTGGCTTGGCCCCCACAGATGTGAAGAG 1706
Db   7673 ATATAAATATAAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAAGAG 7732

Qy   1707 GTACACTACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCTTTGTGCTAGGGTTCTTGGG 1766
Db   7733 ------AAGAGTGGTGCAGAGAGAAAAAAGAGCAGTGGGAATAGGAGCTTTGTTCCTTGG 7786

Qy   1767 TTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGTTGACGCTGACCGCTCAGTC 1826
Db   7787 GTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCAGCCTCAATGACGCTGACGGTACAGGC 7846

Qy   1827 CCGAACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGCTGTTGGACGTGGTCAAGAG 1886
Db   7847 CAGACAATTATTGTCTGGTATAGTGCAGCAGCAGAACAATTTGCTGAGGGCTATTGAGGC 7906

Qy   1887 ACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACAAAGAACCTCCAGACTAGGGTCAC 1946
Db   7907 GCAACAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAATCCT 7966

Qy   1947 TGCCATCGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATGCTTGGGGATGTGCGTTTAG 2006
Db   7967 GGCTGTGGAAAGATACCTAAAGGATCAACAGCTCCTGGGGATTTGGGGTTGCTCTGGAAA 8026

Qy   2007 ACAAGTCTGCCACACTACTGTACC------------ATGGCCAAATGCAAGTCTAACACC 2054
Db   8027 ACTCATTTGCACCACTGCTGTGCCTTGGAATGCTAGTTGGAGTAATAAATCTCTGGAACA 8086

Qy   2055 AAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGGAAGAAAA 2114
Db   8087 GATTTGGAATCACACGACCTGGATGGAGTGGGACAGAGAAATTAACAATTACACAAGCTT 8146

Qy   2115 TATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATGAATTACA 2174
Db   8147 AATACACTCCTTAATTGAAGAATCGCAAAACCAGCAAGAAAAGAATGAACAAGAATTATT 8206

Qy   2175 AAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGATAAAGTA 2234
Db   8207 GGAATTAGATAAATGGGCAAGTTTGTGGAATTGGTTTAACATAACAAATTGGCTGTGGTA 8266

Qy   2235 TATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGATCTATAT 2294
Db   8267 TATAAAATTATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGCTGT 8326

Qy   2295 AGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTCTCTTCCCCACCCTC 2354
Db   8327 ACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCACCATTATCGTTT----------- 8375

Qy   2355 TTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACCAGAGAAGGCAAAGA 2414
Db   8376 ----------CAGACCCACCTCCCAACCCCGAGGGGACCCGACAGGCCCGAAGGAATAGA 8425
```

```
Qy   2415 AAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAATATATTCA 2474
Db   8426 AGAAGAAGGTGGAGAGAGAGACAGAGACAGATCCATTCGATTAGTGAACGGATCCTTGGC 8485

Qy   2475 TTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGCAACTGCAGAACCTT 2534
Db   8486 ACTTATCTGGGACGATCTGCGGAGCCTGTGCCTCTTCAGCTACCACCGCTTGAGAGACTT 8545

Qy   2535 GCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTCTCTGCGACCCT 2591
Db   8546 ACTCTTGATTGTAACGAGGATTGTGGAACTTCTGGGACGCAGGGGGTGGGAAGCCCT 8602


RESULT 7
US-10-003-035-24
; Sequence 24, Application US/10003035
; Patent No. US20020155127A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Danher
;  TITLE OF INVENTION: GENETIC VACCINE AGAINST HUMAN IMMUNODEFICIENCY VIRUS
;  FILE REFERENCE: 22488-712
;  CURRENT APPLICATION NUMBER: US/10/003,035
;  CURRENT FILING DATE:  2001-11-01
;  PRIOR APPLICATION NUMBER: 09/585,599
;  PRIOR FILING DATE: 2000-06-02
;  PRIOR APPLICATION NUMBER: PCT/US01/18238
;  PRIOR FILING DATE: 2001-06-04
;  NUMBER OF SEQ ID NOS: 75
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24
;  LENGTH: 2583
;  TYPE: DNA
;  ORGANISM: Artificial sequence
;  FEATURE:
;  OTHER INFORMATION: Modified Env
US-10-003-035-24

  Query Match             8.6%;  Score 226.2;  DB 9;  Length 2583;
  Best Local Similarity   48.7%;  Pred. No. 2.1e-48;
  Matches  905;  Conservative    0;  Mismatches  883;  Indels   69;  Gaps    8;

Qy    768 TAATGAAAGTAGATGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGA 827
Db    570 TGATACTACCAGCTATACGTTGACAAGTTGTAACACCTCAGTCATTACACAGGCCTGTCC 629

Qy    828 CAAACATTATTGGGATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCT 887
Db    630 AAAGGTATCCTTTGAGCCAATTCCCATACATTATTGTGCCCCGGCTGGTTTTGCGATTCT 689

Qy    888 TAGATGTAATGACACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTC 947
Db    690 AAAATGTAATAATAAGACGTTCAATGGAACAGGACC---ATGTACAAATGTCAGCACAGT 746

Qy    948 TTCATGCACAAGGATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAG 1007
Db    747 ACAATGTACACATGGAATTAGGCCAGTAGTATCAACTCAACTGCTGTTAAATGGCAGTCT 806

Qy   1008 AGCAGAAAATAGAACTTATATT---TACTGGCATGGTAGGGATAATAGGACTATAATTAG 1064
Db    807 GGCAGAAGAAGAGGTAGTAATTAGATCTGCCAATTTCACAGACAATGCTAAAACCATAAT 866

Qy   1065 TTTAAATAAGTATTATAATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTT 1124
Db    867 AGTACAGCTGAACCAATCTGTAGAAATTAATTGTACAAGACCCAACAACAATACAAGAAA 926

Qy   1125 ACCAGTCACCATTATGTCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGGCCAAA 1184
Db    927 AAGTATCCGTATCCAGAGAGGACCAGGGAGAGCATTTGTTACAATAGGAAAAATAGGAAA 986

Qy   1185 GCAGGCATGGTGTTGGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGAC 1244
Db    987 TATGAGACAAGCACATTGTAACATTAGTAGAGCAAAATGGAATAACACTTTAAAACAGAT 1046

Qy   1245 CATTGTCAAACATCCCAGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGC 1304
Db   1047 AGATAGCAAATTAAGAGAACAATTTGGAAATAATAAAAC------AATAATCTTTAAGCA 1100

Qy   1305 TCCTGGAGGAGGAGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCT 1364
Db   1101 GTCCTCAGGAGGGGACCCAGAAATTGTAACGCACAGTTTTAATTGTGGAGGGGAATTTTT 1160

Qy   1365 CTACTGTAAAATGAATTGGTTTCTAAATTGGGTAGAAGATA---------GGAATACAGC 1415
Db   1161 CTACTGTAATTCAACACAACTGTTTAATAGTACTTGGTTTAATAGTACTTGGAGTACTAA 1220

Qy   1416 TAACCAGAAGCCAAAGGAACAGCATAAAAGGAATTACGTGCCATGTCATATTAGACAAAT 1475
Db   1221 AGGGTCAAATAACACTGAAGGAAGTGACACAATCACCCTCCCATGCAGAATAAAACAAAT 1280

Qy   1476 AATCAACACTTGGCATAAAGTAGGCAAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCT 1535
Db   1281 TATAAACATGTGGCAGGAAGTAGGAAAAGCAATGTATGCCCCTCCCATCAGTGGACAAAT 1340

Qy   1536 CACGTGTAACTCCACAGTGACCAGTCTCATAGCAAACATAGATTGGATTGATGGAAACCA 1595
Db   1341 TAGATGTTCATCAAATATTACAGGGCTGCTATTAACAAGAGATGGTGGTAATAGCAACAA 1400

Qy   1596 AACTAATATCACCATGAGTGCAGAGGTGGCAGAACTG---------TATCGATTGGAATT 1646
Db   1401 TGAGTCCGAGATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATT 1460

Qy   1647 GGGAGATTATAAATTAGTAGAGATCACTCCAATTGGCTTGGCCCCCACAGATGTGAAGAG 1706
Db   1461 ATATAAATATAAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAAGAG 1520

Qy   1707 GTACACTACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCTTTGTGCTAGGGTTCTTGGG 1766
Db   1521 ------AAGAGTGGTGCAGAGAGAAAAAAGAGCAGTGGGAATAGGAGCTTTGTTCCTTGG 1574

Qy   1767 TTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGTTGACGCTGACCGCTCAGTC 1826
Db   1575 GTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCAGCGTCAATGACGCTGACGGTACAGGC 1634

Qy   1827 CCGAACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGCTGTTGGACGTGGTCAAGAG 1886
Db   1635 CAGACAATTATTGTCTGGTATAGTGCAGCAGCAGAACAATTTGCTGAGGGCTATTGAGGC 1694

Qy   1887 ACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACAAAGAACCTCCAGACTAGGGTCAC 1946
Db   1695 GCAACAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAATCCT 1754

Qy   1947 TGCCATCGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATGCTTGGGGATGTGCGTTTAG 2006
Db   1755 GGCTGTGGAAAGATACCTAAAGGATCAACAGCTCCTGGGGATTTGGGGTTGCTCTGGAAA 1814

Qy   2007 ACAAGTCTGCCACACTACTGTACC------------ATGGCCAAATGCAAGTCTAACACC 2054
Db   1815 ACTCATTTGCACCACTGCTGTGCCTTGGAATGCTAGTTGGAGTAATAAATCTCTGGAACA 1874

Qy   2055 AAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGGAAGAAAA 2114
Db   1875 GATTTGGAATAACATGACCTGGATGGAGTGGGACAGAGAAATTAACAATTACACAAGCTT 1934

Qy   2115 TATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATGAATTACA 2174
Db   1935 AATACACTCCTTAATTGAAGAATCGCAAAACCAGCAAGAAAAGAATGAACAAGAATTATT 1994

Qy   2175 AAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGATAAAGTA 2234
Db   1995 GGAATTAGATAAATGGGCAAGTTTGTGGAATTGGTTTAACATAACAAATTGGCTGTGGTA 2054

Qy   2235 TATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGATCTATAT 2294
Db   2055 TATAAAATTATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGCTGT 2114

Qy   2295 AGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTCTCTTCCCCACCCTC 2354
```

```
Db   2115 ACTTTCTGTAGTGAATAGAGTTAGGCAGGGATATTCACCATTATCGTTT---------- 2163

Qy   2355 TTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACCAGAGAAGGCAAAGA 2414

Db   2164 ----------CAGACCCACCTCCCAATCCCGAGGGGACCCGACAGGCCCGAAGGAATAGA 2213

Qy   2415 AAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAATATATTCA 2474

Db   2214 AGAAGAAGGTGGAGAGAGAGACAGAGACAGATCCATTCGATTAGTGAACGGATCCTTAGC 2273

Qy   2475 TTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGCAACTGCAGAACCTT 2534

Db   2274 ACTTATCTGGGACGATCTGCGGAGCCTGTGCCTCTTCAGCTACCACCGCTTGAGAGACTT 2333

Qy   2535 GCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTCTCTGCGACCCT 2591

Db   2334 ACTCTTGATTGTAACGAGGATTGTGGAACTTCTGGGACGCAGGGGGTGGGAAGCCCT 2390


RESULT 8
US-10-003-035-14
; Sequence 14, Application US/10003035
; Patent No. US20020155127A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Danher
;  TITLE OF INVENTION: GENETIC VACCINE AGAINST HUMAN IMMUNODEFICIENCY VIRUS
;  FILE REFERENCE: 22488-712
;  CURRENT APPLICATION NUMBER: US/10/003,035
;  CURRENT FILING DATE:  2001-11-01
;  PRIOR APPLICATION NUMBER: 09/585,599
;  PRIOR FILING DATE: 2000-06-02
;  PRIOR APPLICATION NUMBER: PCT/US01/18238
;  PRIOR FILING DATE: 2001-06-04
;  NUMBER OF SEQ ID NOS: 75
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
;   LENGTH: 3157
;   TYPE: DNA
;   ORGANISM: Artificial sequence
;   FEATURE:
;   OTHER INFORMATION: Env/Tat/Rev
US-10-003-035-14

  Query Match             8.6%;  Score 226.2;  DB 9;  Length 3157;
  Best Local Similarity   48.7%;  Pred. No. 2.4e-48;
  Matches  905;  Conservative    0;  Mismatches  883;  Indels   69;  Gaps    8;

Qy    768 TAATGAAAGTAGATGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGA 827

Db   1037 TGATACTACCAGCTATACGTTGACAAGTTGTAACACCTCAGTCATTACACAGGCCTGTCC 1096

Qy    828 CAAACATTATTGGGATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCT 887

Db   1097 AAAGGTATCCTTTGAGCCAATTCCCATACATTATTGTGCCCCGGCTGGTTTTGCGATTCT 1156

Qy    888 TAGATGTAATGACACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTC 947

Db   1157 AAAATGTAATAATAAGACGTTCAATGGAACAGGACC---ATGTACAAATGTCAGCACAGT 1213

Qy    948 TTCATGCACAAGGATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAG 1007

Db   1214 ACAATGTACACATGGAATTAGGCCAGTAGTATCAACTCAACTGCTGTTAAATGGCAGTCT 1273

Qy   1008 AGCAGAAAATAGAACTTATATT---TACTGGCATGGTAGGGATAATAGGACTATAATTAG 1064

Db   1274 GGCAGAAGAAGAGGTAGTAATTAGATCTGCCAATTTCACAGACAATGCTAAAACCATAAT 1333

Qy   1065 TTTAAATAAGTATTATAATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTT 1124

Db   1334 AGTACAGCTGAACCAATCTGTAGAAATTAATTGTACAAGACCCAACAACAATACAAGAAA 1393

Qy   1125 ACCAGTCACCATTATGTCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGGCCAAA 1184

Db   1394 AAGTATCCGTATCCAGAGAGGACCAGGGAGAGCATTTGTTACAATAGGAAAAATAGGAAA 1453

Qy   1185 GCAGGCATGGTGTTGGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGAC 1244

Db   1454 TATGAGACAAGCACATTGTAACATTAGTAGAGCAAAATGGAATAACACTTTAAAACAGAT 1513

Qy   1245 CATTGTCAAACATCCCAGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGC 1304

Db   1514 AGATAGCAAATTAAGAGAACAATTTGGAAATAATAAAAC------AATAATCTTTAAGCA 1567

Qy   1305 TCCTGGAGGAGGAGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCT 1364

Db   1568 GTCCTCAGGAGGGGACCCAGAAATTGTAACGCACAGTTTTAATTGTGGAGGGGAATTTTT 1627

Qy   1365 CTACTGTAAAATGAATTGGTTTCTAAATTGGGTAGAAGATA---------GGAATACAGC 1415

Db   1628 CTACTGTAATTCAACACAACTGTTTAATAGTACTTGGTTTAATAGTACTTGGAGTACTAA 1687

Qy   1416 TAACCAGAAGCCAAAGGAACAGCATAAAAGGAATTACGTGCCATGTCATATTAGACAAAT 1475

Db   1688 AGGGTCAAATAACACTGAAGGAAGTGACACAATCACCCTCCCATGCAGAATAAAACAAAT 1747

Qy   1476 AATCAACACTTGGCATAAAGTAGGCAAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCT 1535

Db   1748 TATAAACATGTGGCAGGAAGTAGGAAAAGCAATGTATGCCCCTCCCATCAGTGGACAAAT 1807

Qy   1536 CACGTGTAACTCCACAGTGACCAGTCTCATAGCAAACATAGATTGGATTGATGGAAACCA 1595

Db   1808 TAGATGTTCATCAAATATTACAGGGCTGCTATTAACAAGAGATGGTGGTAATAGCAACAA 1867

Qy   1596 AACTAATATCACCATGAGTGCAGAGGTGGCAGAACTG---------TATCGATTGGAATT 1646

Db   1868 TGAGTCCGAGATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATT 1927

Qy   1647 GGGAGATTATAAATTAGTAGAGATCACTCCAATTGGCTTGGCCCCCACAGATGTGAAGAG 1706

Db   1928 ATATAAATATAAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAAGAG 1987

Qy   1707 GTACACTACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCTTTGTGCTAGGGTTCTTGGG 1766

Db   1988 ------AAGAGTGGTGCAGAGAGAAAAAAGAGCAGTGGGAATAGGAGCTTTGTTCCTTGG 2041

Qy   1767 TTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGTTGACGCTGACCGCTCAGTC 1826

Db   2042 GTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCAGCGTCAATGACGCTGACGGTACAGGC 2101

Qy   1827 CCGAACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGCTGTTGGACGTGGTCAAGAG 1886

Db   2102 CAGACAATTATTGTCTGGTATAGTGCAGCAGCAGAACAATTTGCTGAGGGCTATTGAGGC 2161

Qy   1887 ACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACAAAGAACCTCCAGACTAGGGTCAC 1946

Db   2162 GCAACAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAATCCT 2221

Qy   1947 TGCCATCGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATGCTTGGGGATGTGCGTTTAG 2006

Db   2222 GGCTGTGGAAAGATACCTAAAGGATCAACAGCTCCTGGGGATTTGGGGTTGCTCTGGAAA 2281

Qy   2007 ACAAGTCTGCCACACTACTGTACC------------ATGGCCAAATGCAAGTCTAACACC 2054

Db   2282 ACTCATTTGCACCACTGCTGTGCCTTGGAATGCTAGTTGGAGTAATAAATCTCTGGAACA 2341

Qy   2055 AAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGGAAGAAAA 2114

Db   2342 GATTTGGAATAACATGACCTGGATGGAGTGGGACAGAGAAATTAACAATTACACAAGCTT 2401

Qy   2115 TATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATGAATTACA 2174

Db   2402 AATACACTCCTTAATTGAAGAATCGCAAAACCAGCAAGAAAAGAATGAACAAGAATTATT 2461

Qy   2175 AAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGATAAAGTA 2234
```

```
Db   2462 GGAATTAGATAAATGGGCAAGTTTGTGGAATTGGTTTAACATAACAAATTGGCTGTGGTA 2521

Qy   2235 TATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGATCTATAT 2294

Db   2522 TATAAAATTATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGCTGT 2581

Qy   2295 AGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTCTCTTCCCCACCCTC 2354

Db   2582 ACTTTCTGTAGTGAATAGAGTTAGGCAGGGATATTCACCATTATCGTTT---------- 2630

Qy   2355 TTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACCAGAGAAGGCAAAGA 2414

Db   2631 ----------CAGACCCACCTCCCAATCCCGAGGGGACCCGACAGGCCCGAAGGAATAGA 2680

Qy   2415 AAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAATATATTCA 2474

Db   2681 AGAAGAAGGTGGAGAGAGAGACAGAGACAGATCCATTCGATTAGTGAACGGATCCTTAGC 2740

Qy   2475 TTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGCAACTGCAGAACCTT 2534

Db   2741 ACTTATCTGGGACGATCTGCGGAGCCTGTGCCTCTTCAGCTACCACCGCTTGAGAGACTT 2800

Qy   2535 GCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTCTCTGCGACCCT 2591

Db   2801 ACTCTTGATTGTAACGAGGATTGTGGAACTTCTGGGACGCAGGGGGTGGGAAGCCCT 2857


RESULT 9
US-09-943-286-3
; Sequence 3, Application US/09943286
; Patent No. US20020106668A1
; GENERAL INFORMATION:
;  APPLICANT: Nunomura, Kiyotada
;  TITLE OF INVENTION: POLYNUCLEOTIDE AMPLIFICATION METHOD
;  FILE REFERENCE: GP104-02.UT
;  CURRENT APPLICATION NUMBER: US/09/943,286
;  CURRENT FILING DATE:  2001-08-30
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
;   LENGTH: 8933
;   TYPE: RNA
;   ORGANISM: Human Immunodeficiency Virus
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)...(8933)
;   OTHER INFORMATION: Sequence of transcripts produced from the BH10
;   OTHER INFORMATION: plasmid.
US-09-943-286-3

  Query Match             8.6%;  Score 226.2;  DB 10;  Length 8933;
  Best Local Similarity   35.9%;  Pred. No. 4e-48;
  Matches  667;  Conservative  238;  Mismatches  883;  Indels   69;  Gaps    8;

Qy    768 TAATGAAAGTAGATGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGA 827

Db   6138 UGAUACUACCAGCUAUACGUUGACAAGUUGUAACACCUCAGUCAUUACACAGGCCUGUCC 6197

Qy    828 CAAACATTATTGGGATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCT 887

Db   6198 AAAGGUAUCCUUUGAGCCAAUUCCCAUACAUUAUUGUGCCCCGGCUGGUUUUGCGAUUCU 6257

Qy    888 TAGATGTAATGACACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTC 947

Db   6258 AAAAUGUAAUAAUAAGACGUUCAAUGGAACAGGACC---AUGUACAAAUGUCAGCACAGU 6314

Qy    948 TTCATGCACAAGGATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAG 1007

Db   6315 ACAAUGUACACAUGGAAUUAGGCCAGUAGUAUCAACUCAACUGCUGUUAAAUGGCAGUCU 6374

Qy   1008 AGCAGAAAATAGAACTTATATT---TACTGGCATGGTAGGGATAATAGGACTATAATTAG 1064

Db   6375 GGCAGAAGAAGAGGUAGUAAUUAGAUCUGCCAAUUUCACAGACAAUGCUAAAACCAUAAU 6434

Qy   1065 TTTAAATAAGTATTATAATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTT 1124

Db   6435 AGUACAGCUGAACCAAUCUGUAGAAAUUAAUUGUACAAGACCCAACAACAAUACAAGAAA 6494

Qy   1125 ACCAGTCACCATTATGTCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGGCCAAA 1184

Db   6495 AAGUAUCCGUAUCCAGAGAGGACCAGGGAGAGCAUUUGUUACAAUAGGAAAAAUAGGAAA 6554

Qy   1185 GCAGGCATGGTGTTGGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGAC 1244

Db   6555 UAUGAGACAAGCACAUUGUAACAUUAGUAGAGCAAAAUGGAAUAACACUUUAAAACAGAU 6614

Qy   1245 CATTGTCAAACATCCCAGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGC 1304

Db   6615 AGAUAGCAAAUUAAGAGAACAAUUUGGAAAUAAUAAAAC------AAUAAUCUUUAAGCA 6668

Qy   1305 TCCTGGAGGAGGAGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCT 1364

Db   6669 GUCCUCAGGAGGGGACCCAGAAAUUGUAACGCACAGUUUUAAUUGUGGAGGGGAAUUUUU 6728

Qy   1365 CTACTGTAAAATGAATTGGTTTCTAAATTGGGTAGAAGATA---------GGAATACAGC 1415

Db   6729 CUACUGUAAUUCAACACAACUGUUUAAUAGUACUUGGUUUAAUAGUACUUGGAGUACUAA 6788

Qy   1416 TAACCAGAAGCCAAAGGAACAGCATAAAAGGAATTACGTGCCATGTCATATTAGACAAAT 1475

Db   6789 AGGGUCAAAUAACACUGAAGGAAGUGACACAAUCACCCUCCCAUGCAGAAUAAAACAAAU 6848

Qy   1476 AATCAACACTTGGCATAAAGTAGGCAAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCT 1535

Db   6849 UAUAAACAUGUGGCAGGAAGUAGGAAAAGCAAUGUAUGCCCCUCCCAUCAGUGGACAAAU 6908

Qy   1536 CACGTGTAACTCCACAGTGACCAGTCTCATAGCAAACATAGATTGGATTGATGGAAACCA 1595

Db   6909 UAGAUGUUCAUCAAAUAUUACAGGGCUGCUAUUAACAAGAGAUGGUGGUAAUAGCAACAA 6968

Qy   1596 AACTAATATCACCATGAGTGCAGAGGTGGCAGAACTG---------TATCGATTGGAATT 1646

Db   6969 UGAGUCCGAGAUCUUCAGACCUGGAGGAGGAGAUAUGAGGGACAAUUGGAGAAGUGAAUU 7028

Qy   1647 GGGAGATTATAAATTAGTAGAGATCACTCCAATTGGCTTGGCCCCCACAGATGTGAAGAG 1706

Db   7029 AUAUAAAUAUAAAGUAGUAAAAAUUGAACCAUUAGGAGUAGCACCCACCAAGGCAAAGAG 7088

Qy   1707 GTACACTACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCTTTGTGCTAGGGTTCTTGGG 1766

Db   7089 ------AAGAGUGGUGCAGAGAGAAAAAAGAGCAGUGGGAAUAGGAGCUUUGUUCCUUGG 7142

Qy   1767 TTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGTTGACGCTGACCGCTCAGTC 1826

Db   7143 GUUCUUGGGAGCAGCAGGAAGCACUAUGGGCGCAGCGUCAAUGACGCUGACGGUACAGGC 7202

Qy   1827 CCGAACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGCTGTTGGACGTGGTCAAGAG 1886

Db   7203 CAGACAAUUAUUGUCUGGUAUAGUGCAGCAGCAGAACAAUUUGCUGAGGGCUAUUGAGGC 7262

Qy   1887 ACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACAAAGAACCTCCAGACTAGGGTCAC 1946

Db   7263 GCAACAGCAUCUGUUGCAACUCACAGUCUGGGGCAUCAAGCAGCUCCAGGCAAGAAUCCU 7322

Qy   1947 TGCCATCGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATGCTTGGGGATGTGCGTTTAG 2006

Db   7323 GGCUGUGGAAAGAUACCUAAAGGAUCAACAGCUCCUGGGGAUUUGGGGUUGCUCUGGAAA 7382

Qy   2007 ACAAGTCTGCCACACTACTGTACC------------ATGGCCAAATGCAAGTCTAACACC 2054

Db   7383 ACUCAUUUGCACCACUGCUGUGCCUUGGAAUGCUAGUUGGAGUAAUAAAUCUCUGGAACA 7442

Qy   2055 AAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGGAAGAAAA 2114

Db   7443 GAUUUGGAAUAACAUGACCUGGAUGGAGUGGGACAGAGAAAUUAACAAUUACACAAGCUU 7502
```

```
Qy    2115 TATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATGAATTACA 2174
Db    7503 AAUACACUCCUUAAUUGAAGAAUCGCAAAACCAGCAAGAAAAGAAUGAACAAGAAUUAUU 7562

Qy    2175 AAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGATAAAGTA 2234
Db    7563 GGAAUUAGAUAAAUGGGCAAGUUUGUGGAAUUGGUUUAACAUAACAAAUUGGCUGUGGUA 7622

Qy    2235 TATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGATCTATAT 2294
Db    7623 UAUAAAAUUAUUCAUAAUGAUAGUAGGAGGCUUGGUAGGUUUAAGAAUAGUUUUUGCUGU 7682

Qy    2295 AGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTCTCTTCCCCACCCTC 2354
Db    7683 ACUUUCUGUAGUGAAUAGAGUUAGGCAGGGAUAUUCACCAUUAUCGUUU---------- 7731

Qy    2355 TTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACCAGAGAAGGCAAAGA 2414
Db    7732 ----------CAGACCCACCUCCCAAUCCCGAGGGGACCCGACAGGCCCGAAGGAAUAGA 7781

Qy    2415 AAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAATATATTCA 2474
Db    7782 AGAAGAAGGUGGAGAGAGAGACAGAGACAGAUCCAUUCGAUUAGUGAACGGAUCCUUAGC 7841

Qy    2475 TTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGCAACTGCAGAACCTT 2534
Db    7842 ACUUAUCUGGGACGAUCUGCGGAGCCUGUGCCUCUUCAGCUACCACCGCUUGAGAGACUU 7901

Qy    2535 GCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTCTCTGCGACCCT 2591
Db    7902 ACUCUUGAUUGUAACGAGGAUUGUGGAACUUCUGGGACGCAGGGGGUGGGAAGCCCU 7958


RESULT 10
US-09-943-286-4
; Sequence 4, Application US/09943286
; Patent No. US20020106668A1
; GENERAL INFORMATION:
;  APPLICANT: Nunomura, Kiyotada
;  TITLE OF INVENTION: POLYNUCLEOTIDE AMPLIFICATION METHOD
;  FILE REFERENCE: GP104-02.UT
;  CURRENT APPLICATION NUMBER: US/09/943,286
;  CURRENT FILING DATE:  2001-08-30
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
;   LENGTH: 8933
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Sequence of the IAC-Asrc pseudo target
;   NAME/KEY: mutation
;   LOCATION: (4135)...(4155)
;   OTHER INFORMATION: Mutated positions: 4135, 4140-1, 4145, 4150,
;   OTHER INFORMATION: 4152-3, 4155
US-09-943-286-4

  Query Match             8.6%;  Score 226.2;  DB 10;  Length 8933;
  Best Local Similarity   35.9%;  Pred. No. 4e-48;
  Matches  667;  Conservative  238;  Mismatches  883;  Indels   69;  Gaps    8;

Qy     768 TAATGAAAGTAGATGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGA 827
Db    6138 UGAUACUACCAGCUAUACGUUGACAAGUUGUAACACCUCAGUCAUUACACAGGCCUGUCC 6197

Qy     828 CAAACATTATTGGGATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCT 887
Db    6198 AAAGGUAUCCUUUGAGCCAAUUCCCAUACAUUAUUGUGCCCCGGCUGGUUUUGCGAUUCU 6257

Qy     888 TAGATGTAATGACACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTC 947
Db    6258 AAAAUGUAAUAAUAAGACGUUCAAUGGAACAGGACC---AUGUACAAAUGUCAGCACAGU 6314

Qy     948 TTCATGCACAAGGATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAG 1007
Db    6315 ACAAUGUACACAUGGAAUUAGGCCAGUAGUAUCAACUCAACUGCUGUUAAAUGGCAGUCU 6374

Qy    1008 AGCAGAAAATAGAACTTATATT---TACTGGCATGGTAGGGATAATAGGACTATAATTAG 1064
Db    6375 GGCAGAAGAAGAGGUAGUAAUUAGAUCUGCCAAUUUCACAGACAAUGCUAAAACCAUAAU 6434

Qy    1065 TTTAAATAAGTATTATAATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTT 1124
Db    6435 AGUACAGCUGAACCAAUCUGUAGAAAUUAAUUGUACAAGACCCAACAACAAUACAAGAAA 6494

Qy    1125 ACCAGTCACCATTATGTCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGGCCAAA 1184
Db    6495 AAGUAUCCGUAUCCAGAGAGGACCAGGGAGAGCAUUUGUUACAAUAGGAAAAAUAGGAAA 6554

Qy    1185 GCAGGCATGGTGTTGGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGAC 1244
Db    6555 UAUGAGACAAGCACAUUGUAACAUUAGUAGAGCAAAAUGGAAUAACACUUUAAAACAGAU 6614

Qy    1245 CATTGTCAAACATCCCAGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGC 1304
Db    6615 AGAUAGCAAAUUAAGAGAACAAUUUGGAAAUAAUAAAAC------AAUAAUCUUUAAGCA 6668

Qy    1305 TCCTGGAGGAGGAGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCT 1364
Db    6669 GUCCUCAGGAGGGGACCCAGAAAUUGUAACGCACAGUUUUAAUUGUGGAGGGGAAUUUUU 6728

Qy    1365 CTACTGTAAAATGAATTGGTTTCTAAATTGGGTAGAAGATA---------GGAATACAGC 1415
Db    6729 CUACUGUAAUUCAACACAACUGUUUAAUAGUACUUGGUUUAAUAGUACUUGGAGUACUAA 6788

Qy    1416 TAACCAGAAGCCAAAGGAACAGCATAAAAGGAATTACGTGCCATGTCATATTAGACAAAT 1475
Db    6789 AGGGUCAAAUAACACUGAAGGAAGUGACACAAUCACCCUCCCAUGCAGAAUAAAACAAAU 6848

Qy    1476 AATCAACACTTGGCATAAAGTAGGCAAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCT 1535
Db    6849 UAUAAACAUGUGGCAGGAAGUAGGAAAAGCAAUGUAUGCCCCUCCCAUCAGUGGACAAAU 6908

Qy    1536 CACGTGTAACTCCACAGTGACCAGTCTCATAGCAAACATAGATTGGATTGATGGAAACCA 1595
Db    6909 UAGAUGUUCAUCAAAUAUUACAGGGCUGCUAUUAACAAGAGAUGGUGGUAAUAGCAACAA 6968

Qy    1596 AACTAATATCACCATGAGTGCAGAGGTGGCAGAACTG---------TATCGATTGGAATT 1646
Db    6969 UGAGUCCGAGAUCUUCAGACCUGGAGGAGGAGAUAUGAGGGACAAUUGGAGAAGUGAAUU 7028

Qy    1647 GGGAGATTATAAATTAGTAGAGATCACTCCAATTGGCTTGGCCCCCACAGATGTGAAGAG 1706
Db    7029 AUAUAAAUAUAAAGUAGUAAAAAUUGAACCAUUAGGAGUAGCACCCACCAAGGCAAAGAG 7088

Qy    1707 GTACACTACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCTTTGTGCTAGGGTTCTTGGG 1766
Db    7089 ------AAGAGUGGUGCAGAGAGAAAAAAGAGCAGUGGGAAUAGGAGCUUUGUUCCUUGG 7142

Qy    1767 TTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGTTGACGCTGACCGCTCAGTC 1826
Db    7143 GUUCUUGGGAGCAGCAGGAAGCACUAUGGGCGCAGCGUCAAUGACGCUGACGGUACAGGC 7202

Qy    1827 CCGAACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGCTGTTGGACGTGGTCAAGAG 1886
Db    7203 CAGACAAUUAUUGUCUGGUAUAGUGCAGCAGCAGAACAAUUUGCUGAGGGCUAUUGAGGC 7262

Qy    1887 ACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACAAAGAACCTCCAGACTAGGGTCAC 1946
Db    7263 GCAACAGCAUCUGUUGCAACUCACAGUCUGGGGCAUCAAGCAGCUCCAGGCAAGAAUCCU 7322

Qy    1947 TGCCATCGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATGCTTGGGGATGTGCGTTTAG 2006
Db    7323 GGCUGUGGAAAGAUACCUAAAGGAUCAACAGCUCCUGGGGAUUUGGGGUUGCUCUGGAAA 7382

Qy    2007 ACAAGTCTGCCACACTACTGTACC------------ATGGCCAAATGCAAGTCTAACACC 2054
```

```
Db   7383 ACUCAUUUGCACCACUGCUGUGCCUUGGAAUGCUAGUUGGAGUAAUAAAUCUCUGGAACA 7442

Qy   2055 AAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGGAAGAAAA 2114
Db   7443 GAUUUGGAAUAACAUGACCUGGAUGGAGUGGGACAGAGAAAUUAACAAUUACACAAGCUU 7502

Qy   2115 TATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATGAATTACA 2174
Db   7503 AAUACACUCCUUAAUUGAAGAAUCGCAAAACCAGCAAGAAAAGAAUGAACAAGAAUUAUU 7562

Qy   2175 AAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGATAAAGTA 2234
Db   7563 GGAAUUAGAUAAAUGGGCAAGUUUGUGGAAUUGGUUUAACAUAACAAAUUGGCUGUGGUA 7622

Qy   2235 TATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGATCTATAT 2294
Db   7623 UAUAAAAUUAUUCAUAAUGAUAGUAGGAGGCUUGGUAGGUUUAAGAAUAGUUUUUGCUGU 7682

Qy   2295 AGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTCTCTTCCCCACCCTC 2354
Db   7683 ACUUUCUGUAGUGAAUAGAGUUAGGCAGGGAUAUUCACCAUUAUCGUUU---------- 7731

Qy   2355 TTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACCAGAGAAGGCAAAGA 2414
Db   7732 ----------CAGACCCACCUCCCAAUCCCGAGGGGACCCGACAGGCCCGAAGGAAUAGA 7781

Qy   2415 AAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAATATATTCA 2474
Db   7782 AGAAGAAGGUGGAGAGAGAGACAGAGACAGAUCCAUUCGAUUAGUGAACGGAUCCUUAGC 7841

Qy   2475 TTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGCAACTGCAGAACCTT 2534
Db   7842 ACUUAUCUGGGACGAUCUGCGGAGCCUGUGCCUCUUCAGCUACCACCGCUUGAGAGACUU 7901

Qy   2535 GCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTCTCTGCGACCCT 2591
Db   7902 ACUCUUGAUUGUAACGAGGAUUGUGGAACUUCUGGGACGCAGGGGGUGGGAAGCCCU 7958
```

RESULT 11
US-09-943-286-9
; Sequence 9, Application US/09943286
; Patent No. US20020106668A1
; GENERAL INFORMATION:
; APPLICANT: Nunomura, Kiyotada
; TITLE OF INVENTION: POLYNUCLEOTIDE AMPLIFICATION METHOD
; FILE REFERENCE: GP104-02.UT
; CURRENT APPLICATION NUMBER: US/09/943,286
; CURRENT FILING DATE: 2001-08-30
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 8933
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Sequence of the IAC-Bscr pseudo target
; NAME/KEY: mutation
; LOCATION: (4140)...(4159)
; OTHER INFORMATION: Mutated positions: 4140-42, 4145-47, 4152,
; OTHER INFORMATION: 4156-57, 4159
US-09-943-286-9

Query Match 8.6%; Score 226.2; DB 10; Length 8933;
Best Local Similarity 35.9%; Pred. No. 4e-48;
Matches 667; Conservative 238; Mismatches 883; Indels 69; Gaps 8;

```
Qy    768 TAATGAAAGTAGATGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGA 827
Db   6138 UGAUACUACCAGCUAUACGUUGACAAGUUGUAACACCUCAGUCAUUACACAGGCCUGUCC 6197

Qy    828 CAAACATTATTGGGATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCT 887

Db   6198 AAAGGUAUCCUUUGAGCCAAUUCCCAUACAUUAUUGUGCCCCGGCUGGUUUUGCGAUUCU 6257

Qy    888 TAGATGTAATGACACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTC 947
Db   6258 AAAAUGUAAUAAUAAGACGUUCAAUGGAACAGGACC---AUGUACAAAUGUCAGCACAGU 6314

Qy    948 TTCATGCACAAGGATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAG 1007
Db   6315 ACAAUGUACACAUGGAAUUAGGCCAGUAGUAUCAACUCAACUGCUGUUAAAUGGCAGUCU 6374

Qy   1008 AGCAGAAAATAGAACTTATATT---TACTGGCATGGTAGGGATAATAGGACTATAATTAG 1064
Db   6375 GGCAGAAGAAGAGGUAGUAAUUAGAUCUGCCAAUUUCACAGACAAUGCUAAAACCAUAAU 6434

Qy   1065 TTTAAATAAGTATTATAATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTT 1124
Db   6435 AGUACAGCUGAACCAAUCUGUAGAAAUUAAUUGUACAAGACCCAACAACAAUACAAGAAA 6494

Qy   1125 ACCAGTCACCATTATGTCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGGCCAAA 1184
Db   6495 AAGUAUCCGUAUCCAGAGAGGACCAGGGAGAGCAUUUGUUACAAUAGGAAAAAUAGGAAA 6554

Qy   1185 GCAGGCATGGTGTTGGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGAC 1244
Db   6555 UAUGAGACAAGCACAUUGUAACAUUAGUAGAGCAAAAUGGAAUAACACUUUAAAACAGAU 6614

Qy   1245 CATTGTCAAACATCCCAGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGC 1304
Db   6615 AGAUAGCAAAUUAAGAGAACAAUUUGGAAAUAAUAAAAC------AAUAAUCUUUAAGCA 6668

Qy   1305 TCCTGGAGGAGGAGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCT 1364
Db   6669 GUCCUCAGGAGGGGACCCAGAAAUUGUAACGCACAGUUUUAAUUGUGGAGGGGAAUUUUU 6728

Qy   1365 CTACTGTAAAATGAATTGGTTTCTAAATTGGGTAGAAGATA---------GGAATACAGC 1415
Db   6729 CUACUGUAAUUCAACACAACUGUUUAAUAGUACUUGGUUUAAUAGUACUUGGAGUACUAA 6788

Qy   1416 TAACCAGAAGCCAAAGGAACAGCATAAAAGGAATTACGTGCCATGTCATATTAGACAAAT 1475
Db   6789 AGGGUCAAAUAACACUGAAGGAAGUGACACAAUCACCCUCCCAUGCAGAAUAAAACAAAU 6848

Qy   1476 AATCAACACTTGGCATAAAGTAGGCAAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCT 1535
Db   6849 UAUAAACAUGUGGCAGGAAGUAGGAAAAGCAAUGUAUGCCCCUCCCAUCAGUGGACAAAU 6908

Qy   1536 CACGTGTAACTCCACAGTGACCAGTCTCATAGCAAACATAGATTGGATTGATGGAAACCA 1595
Db   6909 UAGAUGUUCAUCAAAUAUUACAGGGCUGCUAUUAACAAGAGAUGGUGGUAAUAGCAACAA 6968

Qy   1596 AACTAATATCACCATGAGTGCAGAGGTGGCAGAACTG---------TATCGATTGGAATT 1646
Db   6969 UGAGUCCGAGAUCUUCAGACCUGGAGGAGGAGAUAUGAGGGACAAUUGGAGAAGUGAAUU 7028

Qy   1647 GGGAGATTATAAATTAGTAGAGATCACTCCAATTGGCTTGGCCCCCACAGATGTGAAGAG 1706
Db   7029 AUAUAAAUAUAAAGUAGUAAAAAUUGAACCAUUAGGAGUAGCACCCACCAAGGCAAAGAG 7088

Qy   1707 GTACACTACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCTTTGTGCTAGGGTTCTTGGG 1766
Db   7089 ------AAGAGUGGUGCAGAGAGAAAAAAGAGCAGUGGGAAUAGGAGCUUUGUUCCUUGG 7142

Qy   1767 TTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGTTGACGCTGACCGCTCAGTC 1826
Db   7143 GUUCUUGGGAGCAGCAGGAAGCACUAUGGGCGCAGCGUCAAUGACGCUGACGGUACAGGC 7202

Qy   1827 CCGAACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGCTGTTGGACGTGGTCAAGAG 1886
Db   7203 CAGACAAUUAUUGUCUGGUAUAGUGCAGCAGCAGAACAAUUUGCUGAGGGCUAUUGAGGC 7262

Qy   1887 ACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACAAAGAACCTCCAGACTAGGGTCAC 1946
```

```
Db   7263 GCAACAGCAUCUGUUGCAACUCACAGUCUGGGGCAUCAAGCAGCUCCAGGCAAGAAUCCU 7322

Qy   1947 TGCCATCGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATGCTTGGGGATGTGCGTTTAG 2006
Db   7323 GGCUGUGGAAAGAUACCUAAAGGAUCAACAGCUCCUGGGGAUUUGGGGUUGCUCUGGAAA 7382

Qy   2007 ACAAGTCTGCCACACTACTGTACC----------ATGGCCAAATGCAAGTCTAACACC 2054
Db   7383 ACUCAUUUGCACCACUGCUGUGCCUUGGAAUGCUAGUUGGAGUAAUAAAUCUCUGGAACA 7442

Qy   2055 AAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGGAAGAAAA 2114
Db   7443 GAUUUGGAAUAACAUGACCUGGAUGGAGUGGGACAGAGAAAUUAACAAUUACACAAGCUU 7502

Qy   2115 TATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATGAATTACA 2174
Db   7503 AAUACACUCCUUAAUUGAAGAAUCGCAAAACCAGCAAGAAAAGAAUGAACAAGAAUUAUU 7562

Qy   2175 AAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGATAAAGTA 2234
Db   7563 GGAAUUAGAUAAAUGGGCAAGUUUGUGGAAUUGGUUUAACAUAACAAAUUGGCUGUGGUA 7622

Qy   2235 TATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGATCTATAT 2294
Db   7623 UAUAAAAUUAUUCAUAAUGAUAGUAGGAGGCUUGGUAGGUUUAAGAAUAGUUUUUGCUGU 7682

Qy   2295 AGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTCTCTTCCCCACCCTC 2354
Db   7683 ACUUUCUGUAGUGAAUAGAGUUAGGCAGGGAUAUUCACCAUUAUCGUUU---------- 7731

Qy   2355 TTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACCAGAGAAGGCAAAGA 2414
Db   7732 ----------CAGACCCACCUCCCAAUCCCGAGGGGACCCGACAGGCCCGAAGGAAUAGA 7781

Qy   2415 AAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAATATATTCA 2474
Db   7782 AGAAGAAGGUGGAGAGAGAGACAGAGACAGAUCCAUUCGAUUAGUGAACGGAUCCUUAGC 7841

Qy   2475 TTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGCAACTGCAGAACCTT 2534
Db   7842 ACUUAUCUGGGACGAUCUGCGGAGCCUGUGCCUCUUCAGCUACCACCGCUUGAGAGACUU 7901

Qy   2535 GCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTCTCTGCGACCCT 2591
Db   7902 ACUCUUGAUUGUAACGAGGAUUGUGGAACUUCUGGGACGCAGGGGGUGGGAAGCCCU 7958


RESULT 12
US-09-798-675-4
; Sequence 4, Application US/09798675
; Patent No. US20020106798A1
; GENERAL INFORMATION:
;  APPLICANT: Emory University
;  TITLE OF INVENTION: HIV VACCINES
;  FILE REFERENCE: E056 2020
;  CURRENT APPLICATION NUMBER: US/09/798,675
;  CURRENT FILING DATE:  2001-12-11
;  PRIOR APPLICATION NUMBER: US 60/186,364
;  PRIOR FILING DATE: 2000-03-02
;  PRIOR APPLICATION NUMBER: US 60/251,083
;  PRIOR FILING DATE: 2000-12-01
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
;   LENGTH: 9544
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Construct of vaccine vector pGA2 and insert JS2 expressing clade
;   OTHER INFORMATION: HIV-1 VL
;   NAME/KEY: misc_feature
;   LOCATION: (106)..(1641)
;   OTHER INFORMATION: encoding inner core protein
;   NAME/KEY: misc_feature
;   LOCATION: (1401)..(3617)
;   OTHER INFORMATION: encoding viral rproteins including reverse transcriptase
;   NAME/KEY: misc-feature
;   LOCATION: (3708)..(6334)
;   OTHER INFORMATION: induces high level of virus particles
;   NAME/KEY: misc-feature
;   LOCATION: (3847)..(6518)
;   OTHER INFORMATION: encodes Rev regulating transfer of RNA to cytoplasm
;   NAME/KEY: misc_feature
;   LOCATION: (4102)..(6660)
;   OTHER INFORMATION: encoding ADA Env
;   NAME/KEY: misc_feature
;   LOCATION: (6672)..(9544)
;   OTHER INFORMATION: Vaccine vector pGA2
US-09-798-675-4

  Query Match             8.4%;  Score 222.8;  DB 10;  Length 9544;
  Best Local Similarity   49.0%;  Pred. No. 3.1e-47;
  Matches  966;  Conservative    0;  Mismatches  932;  Indels   74;  Gaps   11;

Qy    651 AGAGCAAATGATAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGA 710
Db   4542 AGGAGAAATAAAAAACTGCTCTTTCAATATCACCACAAGCATAAGAGATAAGGTGAAGAA 4601

Qy    711 GTACAATGAAACTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAATAACACTGGTAA 770
Db   4602 AGACTATGCACTTTTTTATAGACTTGATGTAGTACCAATAGA-----------TAATGA 4649

Qy    771 TGAAAGTAGATGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGACAA 830
Db   4650 TAATACTAGCTATAGGTTGATAAATTGTAATACCTCAACCATTACACAGGCCTGTCCAAA 4709

Qy    831 ACATTATTGGGATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCTTAG 890
Db   4710 GGTATCCTTTGAGCCAATTCCCATACATTATTGTACCCCGGCTGGTTTTGCGATTCTAAA 4769

Qy    891 ATGTAATGACACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTCTTC 950
Db   4770 GTGTAAAGACAAGAAGTTCAATGGAACAGGGCC---ATGTAAAAATGTCAGCACAGTACA 4826

Qy    951 ATGCACAAGGATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAGAGC 1010
Db   4827 ATGTACACATGGAATTAGGCCAGTAGTGTCAACTCAACTGCTGTTAAATGGCAGTCTAGC 4886

Qy   1011 AGAAAATAGAACTTATATTTACT---GGCATGGTAGGGATAATAGGACTATAATTAGTTT 1067
Db   4887 AGAAGAAGAGGTAGTAATTAGATCTAGTAATTTCACAGACAATGCAAAAAACATAATAGT 4946

Qy   1068 AAATAAGTATTATAATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACC 1127
Db   4947 ACAGTTGAAAGAATCTGTAGAAATTAATTGTACAAGACCCAACAACAATACAAGGAAAAG 5006

Qy   1128 AGTCACCATTATGTCTGGATTGGTTTTCCACTCAC------AACCAATCAATGATAGGCC 1181
Db   5007 TATACATATAGGACCAGGAAGAGCATTTTATACAACAGGAGAAATAATAGGAGATATAAG 5066

Qy   1182 AAAGCAGGCATGGTGTTGGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCA 1241
Db   5067 ACAAGCACATTGCAACATTAGTAGAACAAAATGGAATAACACTTTAAATCAAATAGCTAC 5126

Qy   1242 GACCATTGTCAAACATCCCAGGTATACTGGAACTAACAATACTGATAAAATCAATTTGAC 1301
Db   5127 AAAATTAAAAGAACAATTTGGGAATAATAAAAC-AATAGTCTTTAATCAATCCTCAGGAG 5185

Qy   1302 GGCTCCTGGAGGAGGAGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTT 1361
Db   5186 GGGACCCAGAAATTGTAATGCACAGTTTTAATTGTGGAGGGGAATTTTTCTACTGTAATT 5245

Qy   1362 CCTCTACTGTAAAATGAATTGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACCA 1421
Db   5246 CAACACAACT--GTTTAATAGTACTTGGAATTTTAATGGTACTTGGAATTTAACACAATC 5303
```

```
Qy   1422 GAAGCCAAAGGAACAGCATAAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCAA 1481
          |||    |  |||    || | |  |      | ||||||   || | |||||| || ||
Db   5304 GAATGGTACTGAAGGAAATGACACTATCACACTCCCATGTAGAATAAAACAAATTATAAA 5363

Qy   1482 CACTTGGCATAAAGTAGGCAAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCTCACGTG 1541
          |  |||||  ||||||| |||    | |||  |||||  |     ||| |  | |  ||
Db   5364 TATGTGGCAGGAAGTAGGAAAAGCAATGTATGCCCCTCCCATCAGAGGACAAATTAGATG 5423

Qy   1542 TAACTCCACAGTGACCAGTCTCATAGCAAACATAGATTG---------GATTGATGGAAA 1592
            || |   | ||  | || |||  ||  | |||| |          | |  |||
Db   5424 CTCATCAAATATTACAGGGCTAATATTAACAAGAGATGGTGGAACTAACAGTAGTGGGTC 5483

Qy   1593 CCAAACTAATATCACCATGAGTGCAGAGGTGGCAGAACTGTATCGATTGGAATTGGGAGA 1652
          | | |     |   |  | | | |||  ||   ||    |  ||   |||||     |
Db   5484 CGAGATCTTCAGACCTGGGGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAA 5543

Qy   1653 TTATAAATTAGTAGAGATCACTCCAATTGGCTTGGCCCCCACAGATGTGAAGAGGTACAC 1712
          |||||| |||||  | ||    ||| | ||  | || |||||  | |  || ||    |
Db   5544 ATATAAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAAAAG----AA 5599

Qy   1713 TACTGGTGGCACCTCAAGAAATAAAAGAGGGGTC-TTTGTGCTAGGGTTCTTGGGTTTTC 1771
          | |||||       || ||| |  || |||  |  | |       |||| | || ||
Db   5600 GAGTGGTGCAGAGAGAAAAAAGAGCAGTGGGAACGATAGGAGCTATGTTCCTTGGGTTCT 5659

Qy   1772 TCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGTTGACGCTGACCGCTCAGTCCCGAA 1831
          | | | | |||||     | |||||||| |||||  | ||||||| |   ||| || ||
Db   5660 TGGGAGCAGCAGGAAGCACTATGGGCGCAGCGTCAATAACGCTGACGGTACAGGCCAGAC 5719

Qy   1832 CTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGCTGTTGGACGTGGTCAAGAGACAAC 1891
            ||||||  |||| |||||||| || ||| |  |  || ||    |  |  ||   ||||
Db   5720 TATTATTGTCTGGTATAGTGCAACAGCAGAACAATTTGCTGAGGGCTATTGAGGCGCAAC 5779

Qy   1892 AAGAATTGTTGCGACTGACCGTCTGGGGAACAAAGAACCTCCAGACTAGGGTCACTGCCA 1951
          |  |  |||||| ||| || |||||||| |  ||| | |||||| | || |||   ||
Db   5780 AGCATCTGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAGTCCTGGCTC 5839

Qy   1952 TCGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATGCTTGGGGATGTGCGTTTAGACAAG 2011
          | || |  ||| ||| ||| ||   ||  ||   · |||||| ||  |    | ||
Db   5840 TGGAAAGATACCTAAGGGATCAACAGCTCCTAGGGATTTGGGGTTGCTCTGGAAAACTCA 5899

Qy   2012 TCTGCCACACTACTGTACC------------ATGGCCAAATGCAAGTCTAACACCAAAGT 2059
          |||||  |||| |||| ||            ||| . |||  || |||       |  |
Db   5900 TCTGCACCACTGCTGTGCCTTGGAATGCTAGTTGGAGTAATAAAACTCTGGATATGATTT 5959

Qy   2060 GGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGGAAGAAAATATAA 2119
          || | ||   ||| |||   |||||||| || |  | ||    |    ||     |||
Db   5960 GGGATAACATGACCTGGATGGAGTGGGAAAGAGAAATCGAAAATTACACAGGCTTAATAT 6019

Qy   2120 CAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATGAATTACAAAAGT 2179
            || |  | || ||  | || |  |||||||| |||||     | || ||| |   |
Db   6020 ACACCTTAATTGAAGAATCGCAGAACCAACAAGAAAAGAATGAACAAGACTTATTAGCAT 6079

Qy   2180 TGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGATAAAGTATATAC 2239
          |  |||  ||||     ||  | |||||||||||| | |   |||| |   || ||||
Db   6080 TAGATAAGTGGGCAAGTTTGTGGAATTGGTTTGACATATCAAATTGGCTGTGGTGTATAA 6139

Qy   2240 AATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGATCTATATAGTAC 2299
          ||       |   ||||| | |||  | |||   ||||||||||| |   | || |
Db   6140 AAATCTTCATAATGATAGTAGGAGGCTTGATAGGTTTAAGAATAGTTTTTACTGTACTTT 6199

Qy   2300 AAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTCTCTTCCCCACCCTCTTATT 2359
            ||  ||   | |||||||| ||    ||| |||  | |
Db   6200 CTATAGTAAATAGAGTTAGGCAGGGATACTCACCATTGTCATTT---------------- 6243

Qy   2360 TCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACCAGAGAAGGCAAAGAAAGAG 2419
          |||||||| ||| |         || ||   ·  |     ||||| |  ||| ||
Db   6244 -----CAGACCCACCTCCCAGCCCCGAGGGGACCCGACAGGCCCGAAGGAATCGAAGAAG 6298

Qy   2420 ACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAATATATTCATTTCC 2479
          | |||||||  | |   |  |||| ||| ||  |    |  | ||  |     |
Db   6299 AAGGTGGAGACAGAGACAGAGACAGATCCGTGCGATTAGTGGATGGATCCTTAGCACTTA 6358

Qy   2480 TGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGCAACTGCAGAACCTTGCTAT 2539
```

```
          |      | | |||    |||| |      |     | |  |  ||  |||  ||| || |
Db   6359 TCTGGGACGATCTGCGGAGCCTGTGCCTCTTCAGCTACCACCGCTTGAGAGACTTACTCT 6418

Qy   2540 CGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTCTCTGCGACCCT 2591
            ||  |||     |||  |  |||    |     |  |||   |  |   ||||     |
Db   6419 TGATTGTAACGAGGATTGTGGAACTTCTGGGACGCAGGGGGTGGGAAGCCCT 6470


RESULT 13
US-10-003-035-22
; Sequence 22, Application US/10003035
; Patent No. US20020155127A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Danher
;  TITLE OF INVENTION: GENETIC VACCINE AGAINST HUMAN IMMUNODEFICIENCY VIRUS
;  FILE REFERENCE: 22488-712
;  CURRENT APPLICATION NUMBER: US/10/003,035
;  CURRENT FILING DATE:  2001-11-01
;  PRIOR APPLICATION NUMBER: 09/585,599
;  PRIOR FILING DATE: 2000-06-02
;  PRIOR APPLICATION NUMBER: PCT/US01/18238
;  PRIOR FILING DATE: 2001-06-04
;  NUMBER OF SEQ ID NOS: 75
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 22
;   LENGTH: 2950
;   TYPE: DNA
;   ORGANISM: Artificial sequence
;   FEATURE:
;   OTHER INFORMATION: Modified Env/Nef from strain BH10
US-10-003-035-22

  Query Match             8.3%;  Score 220.2;  DB 9;  Length 2950;
  Best Local Similarity   48.9%;  Pred. No. 8e-47;
  Matches  909;  Conservative    0;  Mismatches  873;  Indels   75;  Gaps    9;

Qy    768 TAATGAAAGTAGATGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGA 827
            | ||   |  || | |    |||     |||||||| || || ||   | || | |||
Db    570 TGATACTACCAGCTATACGTTGACAAGTTGTAACACCTCAGTCATTACACAGGCCTGTCC 629

Qy    828 CAAACATTATTGGGATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCT 887
            ||    |  |  || | |||     ||||||||  ||  | |||| |||  ||||
Db    630 AAAGGTATCCTTTGAGCCAATTCCCATACATTATTGTGCCCCGGCTGGTTTTGCGATTCT 689

Qy    888 TAGATGTAATGACACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTC 947
            | |||||||  | |  |  |      ||       ||   |||| | || ||
Db    690 AAAATGTAATAATAAGACGTTCAATGGAACAGGACC---ATGTACAAATGTCAGCACAGT 746

Qy    948 TTCATGCACAAGGATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAG 1007
              ||| |||      ||   | ||     || |||     |   || ||||| | ||
Db    747 ACAATGTACACATGGAATTAGGCCAGTAGTATCAACTCAACTGCTGTTAAATGGCAGTCT 806

Qy   1008 AGCAGAAAATAGAACTTATATT---TACTGGCATGGTAGGGATAATAGGACTATAATTAG 1064
            |||||| |          |||    |       ||   |  || |||    |  |  || ||
Db    807 GGCAGAAGAAGAGGTAGTAATTAGATCTGCCAATTTCACAGACAATGCTAAAACCATAAT 866

Qy   1065 TTTAAATAAGTATTATAATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTT 1124
            || |   | |  |   | ||  ||| || |||| ||||||    || || |||
Db    867 AGTACAGCTGAACCAATCTGTAGAAATTAATTGTACAAGACCCAACAACAATACAAGAAA 926

Qy   1125 ACCAGTCACCATTATGTCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGGCCAAA 1184
            |    ||  ||   | |||    ||||      ||||    | |    |||
Db    927 AAGTATCCGTATCCAGAGAGGACCAGGGAGAGCATTTGTTACAATAGGAAAAATAGGAAA 986

Qy   1185 GCAGGCATGGTGTTGGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGAC 1244
            |  |     | |    |   |    |   |  ||    | || ||||
Db    987 TATGAGACAAGCACATTGTAACATTAGTAGAGCAAAATGGAATAACACTTTAAAACAGAT 1046

Qy   1245 CATTGTCAAACATCCCAGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGC 1304
            |  ||||          |  ||||| ||| || ||      ||| |  || | |
Db    1047 AGATAGCAAATTAAGAGAACAATTTGGAAATAATAAAAC------AATAATCTTTAAGCA 1100

Qy   1305 TCCTGGAGGAGGAGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCT 1364
```

```
Db   1101 GTCCTCAGGAGGGGACCCAGAAATTGTAACGCACAGTTTTAATTGTGGAGGGGAATTTTTT 1160

Qy   1365 CTACTGTAAAATGAATTGGTTTCTAAATTGGGTAGAAGATA---------GGAATACAGC 1415
Db   1161 CTACTGTAATTCAACACAACTGTTTAATAGTACTTGGTTTAATAGTACTTGGAGTACTAA 1220

Qy   1416 TAACCAGAAGCCAAAGGAACAGCATAAAAGGAATTACGTGCCATGTCATATTAGACAAAT 1475
Db   1221 AGGGTCAAATAACACTGAAGGAAGTGACACAATCACCCTCCCATGCAGAATAAAACAAAT 1280

Qy   1476 AATCAACACTTGGCATAAAGTAGGCAAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCT 1535
Db   1281 TATAAACATGTGGCAGGAAGTAGGAAAAGCAATGTATGCCCCTCCCATCAGTGGACAAAT 1340

Qy   1536 CACGTGTAACTCCACAGTGACCAGTCTCATAGCAAACATAGATTGGATTGATGGAAACCA 1595
Db   1341 TAGATGTTCATCAAATATTACAGGGCTGCTATTAACAAGAGATGGTGGTAATAGCAACAA 1400

Qy   1596 AACTAATATCACCATGAGTGCAGAGGTGGCAGAACTG---------TATCGATTGGAATT 1646
Db   1401 TGAGTCCGAGATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATT 1460

Qy   1647 GGGAGATTATAAATTAGTAGAGATCACTCCAATTGGCTTGGCCCCCACAGATGTGAAGAG 1706
Db   1461 ATATAAATATAAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAAGAG 1520

Qy   1707 GTACACTACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCTTTGTGCTAGGGTTCTTGGG 1766
Db   1521 ----AAGAGTGGTGCAGACTAGTGCAGTGGGAATAGGAGCTTT--------GTTCCTTGG 1568

Qy   1767 TTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGTTGACGCTGACCGCTCAGTC 1826
Db   1569 GTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCAGCGTCAATGACGCTGACGGTACAGGC 1628

Qy   1827 CCGAACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGCTGTTGGACGTGGTCAAGAG 1886
Db   1629 CAGACAATTATTGTCTGGTATAGTGCAGCAGCAGAACAATTTGCTGAGGGCTATTGAGGC 1688

Qy   1887 ACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACAAAGAACCTCCAGACTAGGGTCAC 1946
Db   1689 GCAACAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAATCCT 1748

Qy   1947 TGCCATCGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATGCTTGGGGATGTGCGTTTAG 2006
Db   1749 GGCTGTGGAAAGATACCTAAAGGATCAACAGCTCCTGGGGATTTGGGGTTGCTCTGGAAA 1808

Qy   2007 ACAAGTCTGCCACACTACTGTACC----------ATGGCCAAATGCAAGTCTAACACC 2054
Db   1809 ACTCATTTGCACCACTGCTGTGCCTTGGAATGCTAGTTGGAGTAATAAATCTCTGGAACA 1868

Qy   2055 AAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGGAAGAAAA 2114
Db   1869 GATTTGGAATAACATGACCTGGATGGAGTGGGACAGAGAAATTAACAATTACACAAGCTT 1928

Qy   2115 TATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATGAATTACA 2174
Db   1929 AATACACTCCTTAATTGAAGAATCGCAAAACCAGCAAGAAAAGAATGAACAAGAATTATT 1988

Qy   2175 AAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGATAAAGTA 2234
Db   1989 GGAATTAGATAAATGGGCAAGTTTGTGGAATTGGTTTAACATAACAAATTGGCTGTGGTA 2048

Qy   2235 TATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGATCTATAT 2294
Db   2049 TATAAAATTATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGCTGT 2108

Qy   2295 AGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTCTCTTCCCCACCCTC 2354
Db   2109 ACTTTCTGTAGTGAATAGAGTTAGGCAGGGATATTCACCATTATCGTTT---------- 2157

Qy   2355 TTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACCAGAGAAGGCAAAGA 2414
```

```
Db   2158 ----------CAGACCCACCTCCCAATCCCGAGGGGACCCGACAGGCCCGAAGGAATAGA 2207

Qy   2415 AAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAATATATTCA 2474
Db   2208 AGAAGAAGGTGGAGAGAGAGACAGAGACAGATCCATTCGATTAGTGAACGGATCCTTAGC 2267

Qy   2475 TTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGCAACTGCAGAACCTT 2534
Db   2268 ACTTATCTGGGACGATCTGCGGAGCCTGTGCCTCTTCAGCTACCACCGCTTGAGAGACTT 2327

Qy   2535 GCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTCTCTGCGACCCT 2591
Db   2328 ACTCTTGATTGTAACGAGGATTGTGGAACTTCTGGGACGCAGGGGGTGGGAAGCCCT 2384


RESULT 14
US-10-003-035-21
; Sequence 21, Application US/10003035
; Patent No. US20020155127A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Danher
;  TITLE OF INVENTION: GENETIC VACCINE AGAINST HUMAN IMMUNODEFICIENCY VIRUS
;  FILE REFERENCE: 22488-712
;  CURRENT APPLICATION NUMBER: US/10/003,035
;  CURRENT FILING DATE:  2001-11-01
;  PRIOR APPLICATION NUMBER: 09/585,599
;  PRIOR FILING DATE: 2000-06-02
;  PRIOR APPLICATION NUMBER: PCT/US01/18238
;  PRIOR FILING DATE: 2001-06-04
;  NUMBER OF SEQ ID NOS: 75
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21
;   LENGTH: 3417
;   TYPE: DNA
;   ORGANISM: Artificial sequence
;   FEATURE:
;   OTHER INFORMATION: Modified Env/Tat/Rev/Nef from strain BH10
US-10-003-035-21

  Query Match             8.3%;  Score 220.2;  DB 9;  Length 3417;
  Best Local Similarity   48.9%;  Pred. No. 8.6e-47;
  Matches  909;  Conservative    0;  Mismatches  873;  Indels   75;  Gaps    9;

Qy    768 TAATGAAAGTAGATGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGA 827
Db   1037 TGATACTACCAGCTATACGTTGACAAGTTGTAACACCTCAGTCATTACACAGGCCTGTCC 1096

Qy    828 CAAACATTATTGGGATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCT 887
Db   1097 AAAGGTATCCTTTGAGCCAATTCCCATACATTATTGTGCCCCGGCTGGTTTTGCGATTCT 1156

Qy    888 TAGATGTAATGACACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTC 947
Db   1157 AAAATGTAATAATAAGACGTTCAATGGAACAGGACC---ATGTACAAATGTCAGCACAGT 1213

Qy    948 TTCATGCACAAGGATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAG 1007
Db   1214 ACAATGTACACATGGAATTAGGCCAGTAGTATCAACTCAACTGCTGTTAAATGGCAGTCT 1273

Qy   1008 AGCAGAAAATAGAACTTATATT---TACTGGCATGGTAGGGATAATAGGACTATAATTAG 1064
Db   1274 GGCAGAAGAAGAGGTAGTAATTAGATCTGCCAATTTCACAGACAATGCTAAAACCATAAT 1333

Qy   1065 TTTAAATAAGTATTATAATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTT 1124
Db   1334 AGTACAGCTGAACCAATCTGTAGAAATTAATTGTACAAGACCCAACAACAATACAAGAAA 1393

Qy   1125 ACCAGTCACCATTATGTCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGGCCAAA 1184
Db   1394 AAGTATCCGTATCCAGAGAGGACCAGGGAGAGCATTTGTTACAATAGGAAAAATAGGAAA 1453

Qy   1185 GCAGGCATGGTGTTGGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGAC 1244
```

```
Db   1454 TATGAGACAAGCACATTGTAACATTAGTAGAGCAAAATGGAATAACACTTTAAAACAGAT 1513

Qy   1245 CATTGTCAAACATCCCAGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGC 1304

Db   1514 AGATAGCAAATTAAGAGAACAATTTGGAAATAATAAAAC------AATAATCTTTAAGCA 1567

Qy   1305 TCCTGGAGGAGGAGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCT 1364

Db   1568 GTCCTCAGGAGGGGACCCAGAAATTGTAACGCACAGTTTTAATTGTGGAGGGGAATTTTT 1627

Qy   1365 CTACTGTAAAATGAATTGGTTTCTAAATTGGGTAGAAGATA---------GGAATACAGC 1415

Db   1628 CTACTGTAATTCAACACAACTGTTTAATAGTACTTGGTTTAATAGTACTTGGAGTACTAA 1687

Qy   1416 TAACCAGAAGCCAAAGGAACAGCATAAAAGGAATTACGTGCCATGTCATATTAGACAAAT 1475

Db   1688 AGGGTCAAATAACACTGAAGGAAGTGACACAATCACCCTCCCATGCAGAATAAAACAAAT 1747

Qy   1476 AATCAACACTTGGCATAAAGTAGGCAAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCT 1535

Db   1748 TATAAACATGTGGCAGGAAGTAGGAAAAGCAATGTATGCCCCTCCCATCAGTGGACAAAT 1807

Qy   1536 CACGTGTAACTCCACAGTGACCAGTCTCATAGCAAACATAGATTGGATTGATGGAAACCA 1595

Db   1808 TAGATGTTCATCAAATATTACAGGGCTGCTATTAACAAGAGATGGTGGTAATAGCAACAA 1867

Qy   1596 AACTAATATCACCATGAGTGCAGAGGTGGCAGAACTG---------TATCGATTGGAATT 1646

Db   1868 TGAGTCCGAGATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATT 1927

Qy   1647 GGGAGATTATAAATTAGTAGAGATCACTCCAATTGGCTTGGCCCCCACAGATGTGAAGAG 1706

Db   1928 ATATAAATATAAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAAGAG 1987

Qy   1707 GTACACTACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCTTTGTGCTAGGGTTCTTGGG 1766

Db   1988 ----AAGAGTGGTGCAGACTAGTGCAGTGGGAATAGGAGCTTT--------GTTCCTTGG 2035

Qy   1767 TTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGTTGACGCTGACCGCTCAGTC 1826

Db   2036 GTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCAGCGTCAATGACGCTGACGGTACAGGC 2095

Qy   1827 CCGAACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGCTGTTGGACGTGGTCAAGAG 1886

Db   2096 CAGACAATTATTGTCTGGTATAGTGCAGCAGCAGAACAATTTGCTGAGGGCTATTGAGGC 2155

Qy   1887 ACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACAAAGAACCTCCAGACTAGGGTCAC 1946

Db   2156 GCAACAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAATCCT 2215

Qy   1947 TGCCATCGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATGCTTGGGGATGTGCGTTTAG 2006

Db   2216 GGCTGTGGAAAGATACCTAAAGGATCAACAGCTCCTGGGGATTTGGGGTTGCTCTGGAAA 2275

Qy   2007 ACAAGTCTGCCACACTACTGTACC----------ATGGCCAAATGCAAGTCTAACACC 2054

Db   2276 ACTCATTTGCACCACTGCTGTGCCTTGGAATGCTAGTTGGAGTAATAAATCTCTGGAACA 2335

Qy   2055 AAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGGAAGAAAA 2114

Db   2336 GATTTGGAATAACATGACCTGGATGGAGTGGGACAGAGAAATTAACAATTACACAAGCTT 2395

Qy   2115 TATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATGAATTACA 2174

Db   2396 AATACACTCCTTAATTGAAGAATCGCAAAACCAGCAAGAAAAGAATGAACAAGAATTATT 2455

Qy   2175 AAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGATAAAGTA 2234

Db   2456 GGAATTAGATAAATGGGCAAGTTTGTGGAATTGGTTTAACATAACAAATTGGCTGTGGTA 2515

Qy   2235 TATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGATCTATAT 2294

Db   2516 TATAAAATTATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGCTGT 2575

Qy   2295 AGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTCTCTTCCCCACCCTC 2354

Db   2576 ACTTTCTGTAGTGAATAGAGTTAGGCAGGGATATTCACCATTATCGTTT---------- 2624

Qy   2355 TTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACCAGAGAAGGCAAAGA 2414

Db   2625 ----------CAGACCCACCTCCCAATCCCGAGGGGACCCGACAGGCCCGAAGGAATAGA 2674

Qy   2415 AAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAATATATTCA 2474

Db   2675 AGAAGAAGGTGGAGAGAGAGACAGAGACAGATCCATTCGATTAGTGAACGGATCCTTAGC 2734

Qy   2475 TTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTCAGCAACTGCAGAACCTT 2534

Db   2735 ACTTATCTGGGACGATCTGCGGAGCCTGTGCCTCTTCAGCTACCACCGCTTGAGAGACTT 2794

Qy   2535 GCTATCGAGAGTATACCAGATCCTCCAACCAATACTCCAGAGGCTCTCTGCGACCCT 2591

Db   2795 ACTCTTGATTGTAACGAGGATTGTGGAACTTCTGGGACGCAGGGGGTGGGAAGCCCT 2851


RESULT 15
US-10-003-035-16
; Sequence 16, Application US/10003035
; Patent No. US20020155127A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Danher
;  TITLE OF INVENTION: GENETIC VACCINE AGAINST HUMAN IMMUNODEFICIENCY VIRUS
;  FILE REFERENCE: 22488-712
;  CURRENT APPLICATION NUMBER: US/10/003,035
;  CURRENT FILING DATE:  2001-11-01
;  PRIOR APPLICATION NUMBER: 09/585,599
;  PRIOR FILING DATE: 2000-06-02
;  PRIOR APPLICATION NUMBER: PCT/US01/18238
;  PRIOR FILING DATE: 2001-06-04
;  NUMBER OF SEQ ID NOS: 75
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16
;   LENGTH: 2280
;   TYPE: DNA
;   ORGANISM: Artificial sequence
;   FEATURE:
;   OTHER INFORMATION: Modified Env
US-10-003-035-16

  Query Match             8.3%;  Score 218.2;  DB 9;  Length 2280;
  Best Local Similarity   49.2%;  Pred. No. 2.3e-46;
  Matches  843;  Conservative    0;  Mismatches  818;  Indels   54;  Gaps    8;

Qy    768 TAATGAAAGTAGATGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAGTCTTGTGA 827

Db    570 TGATACTACCAGCTATACGTTGACAAGTTGTAACACCTCAGTCATTACACAGGCCTGTCC 629

Qy    828 CAAACATTATTGGGATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTATGCTTTGCT 887

Db    630 AAAGGTATCCTTTGAGCCAATTCCCATACATTATTGTGCCCCGGCTGGTTTTGCGATTCT 689

Qy    888 TAGATGTAATGACACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGGTGGTCTC 947

Db    690 AAAATGTAATAATAAGACGTTCAATGGAACAGGACC---ATGTACAAATGTCAGCACAGT 746

Qy    948 TTCATGCACAAGGATGATGGAGACACAGACTTCTACTTGGTTTGGCTTTAATGGAACTAG 1007

Db    747 ACAATGTACACATGGAATTAGGCCAGTAGTATCAACTCAACTGCTGTTAAATGGCAGTCT 806

Qy   1008 AGCAGAAAATAGAACTTATATT---TACTGGCATGGTAGGGATAATAGGACTATAATTAG 1064

Db    807 GGCAGAAGAAGAGGTAGTAATTAGATCTGCCAATTTCACAGACAATGCTAAAACCATAAT 866

Qy   1065 TTTAAATAAGTATTATAATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTT 1124

Db    867 AGTACAGCTGAACCAATCTGTAGAAATTAATTGTACAAGACCCAACAACAATACAAGAAA 926
```

```
Qy   1125 ACCAGTCACCATTATGTCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGGCCAAA 1184
Db    927 AAGTATCCGTATCCAGAGAGGACCAGGGAGAGCATTTGTTACAATAGGAAAAATAGGAAA 986

Qy   1185 GCAGGCATGGTGTTGGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGAC 1244
Db    987 TATGAGACAAGCACATTGTAACATTAGTAGAGCAAAATGGAATAACACTTTAAAACAGAT 1046

Qy   1245 CATTGTCAAACATCCCAGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGC 1304
Db   1047 AGATAGCAAATTAAGAGAACAATTTGGAAATAATAAAAC------AATAATCTTTAAGCA 1100

Qy   1305 TCCTGGAGGAGGAGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCT 1364
Db   1101 GTCCTCAGGAGGGGACCCAGAAATTGTAACGCACAGTTTTAATTGTGGAGGGGAATTTTT 1160

Qy   1365 CTACTGTAAAATGAATTGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAA 1424
Db   1161 CTACTGTAATTCAACACAACTGTTTAATAGTACTTGGTTTAATAGTACTTGGAGTACTAA 1220

Qy   1425 GCCAAAGGAACAGCATAAAAGGAATTACG---------TGCCATGTCATATTAGACAAAT 1475
Db   1221 AGGGTCAAATAACACTGAAGGAAGTGACACAATCACCCTCCCATGCAGAATAAAACAAAT 1280

Qy   1476 AATCAACACTTGGCATAAAGTAGGCAAAAATGTTTATTTGCCTCCAAGAGAGGGAGACCT 1535
Db   1281 TATAAACATGTGGCAGGAAGTAGGAAAAGCAATGTATGCCCCTCCCATCAGTGGACAAAT 1340

Qy   1536 CACGTGTAACTCCACAGTGACCAGTCTCATAGCAAACATAGATTGGATTGATGGAAACCA 1595
Db   1341 TAGATGTTCATCAAATATTACAGGGCTGCTATTAACAAGAGATGGTGGTAATAGCAACAA 1400

Qy   1596 AACTAATATCACCATGAGTGCAGAGGTGGCAGAACTG---------TATCGATTGGAATT 1646
Db   1401 TGAGTCCGAGATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATT 1460

Qy   1647 GGGAGATTATAAATTAGTAGAGATCACTCCAATTGGCTTGGCCCCCACAGATGTGAAGAG 1706
Db   1461 ATATAAATATAAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAAGAG 1520

Qy   1707 GTACACTACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCTTTGTGCTAGGGTTCTTGGG 1766
Db   1521 ----AAGAGTGGTGCAGACTAGTGCAGTGGGAATAGGAGCTTT--------GTTCCTTGG 1568

Qy   1767 TTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCGTTGACGCTGACCGCTCAGTC 1826
Db   1569 GTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCAGCGTCAATGACGCTGACGGTACAGGC 1628

Qy   1827 CCGAACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGCTGTTGGACGTGGTCAAGAG 1886
Db   1629 CAGACAATTATTGTCTGGTATAGTGCAGCAGCAGAACAATTTGCTGAGGGCTATTGAGGC 1688

Qy   1887 ACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACAAAGAACCTCCAGACTAGGGTCAC 1946
Db   1689 GCAACAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAATCCT 1748

Qy   1947 TGCCATCGAGAAGTACTTAAAGGACCAGGCGCAGCTGAATGCTTGGGGATGTGCGTTTAG 2006
Db   1749 GGCTGTGGAAAGATACCTAAAGGATCAACAGCTCCTGGGGATTTGGGGTTGCTCTGGAAA 1808

Qy   2007 ACAAGTCTGCCACACTACTGTACC------------ATGGCCAAATGCAAGTCTAACACC 2054
Db   1809 ACTCATTTGCACCACTGCTGTGCCTTGGAATGCTAGTTGGAGTAATAAATCTCTGGAACA 1868

Qy   2055 AAAGTGGAACAATGAGACTTGGCAAGAGTGGGAGCGAAAGGTTGACTTCTTGGAAGAAAA 2114
Db   1869 GATTTGGAATAACATGACCTGGATGGAGTGGGACAGAGAAATTAACAATTACACAAGCTT 1928

Qy   2115 TATAACAGCCCTCCTAGAGGAGGCACAAATTCAACAAGAGAAGAACATGTATGAATTACA 2174
Db   1929 AATACACTCCTTAATTGAAGAATCGCAAAACCAGCAAGAAAAGAATGAACAAGAATTATT 1988

Qy   2175 AAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGATAAAGTA 2234
Db   1989 GGAATTAGATAAATGGGCAAGTTTGTGGAATTGGTTTAACATAACAAATTGGCTGTGGTA 2048

Qy   2235 TATACAATATGGAGTTTATATAGTTGTAGGAGTAATACTGTTAAGAATAGTGATCTATAT 2294
Db   2049 TATAAAATTATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGCTGT 2108

Qy   2295 AGTACAAATGCTAGCTAAGTTAAGGCAGGGGTATAGGCCAGTGTTCTCTTCCCCACCCTC 2354
Db   2109 ACTTTCTGTAGTGAATAGAGTTAGGCAGGGATATTCACCATTATCGTTTCAGACCCACCT 2168

Qy   2355 TTATTTCCAGCAGACCCATATCCAACAGGACCCGGCACTGCCAACCAGAGAAGGCAAAGA 2414
Db   2169 CCCAATCCCGAGGGGACCCGACAGGCCCGAAGGAATAGAAGAAGAAGGTGGAGAGAGAGA 2228

Qy   2415 AAGAGACGGTGGAGAAGGCGGTGGCAACAGCTCCT 2449
Db   2229 CAGAGACAGATCCATTCGATTAGTGAACGGATCCT 2263
```

Search completed: November 16, 2002, 13:16:23
Job time : 276.157 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 16, 2002, 07:32:18 ; Search time 2447.32 Seconds
(without alignments)
17477.179 Million cell updates/sec

Title: M33262_COPY_6695_9335
Perfect score: 2641
Sequence: 1 TCTAGAAGCATGCTATAACA.....ACTGACCTACCTACAATATG 2641

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results prédicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	58.4	2.2	1101	17	CNS0039G	AL063921 Drosophil
	2	52.6	2.0	592	9	AU039090	AU039090 AU039090
	3	51.4	1.9	230	9	AU261463	AU261463 AU261463
	4	51.4	1.9	1101	17	CNS0182P	AL108811 Drosophil
c	5	50.8	1.9	1101	17	CNS00KLH	AL077724 Drosophil
	6	50.6	1.9	244	9	AU061377	AU061377 AU061377
	7	50.6	1.9	1101	17	CNS0100X	AL098379 Drosophil
c	8	50.2	1.9	991	17	CNS005IJ	AL059544 Drosophil
	9	50	1.9	1101	17	CNS0106X	AL098595 Drosophil
	10	49.4	1.9	1101	17	CNS00EVL	AL069706 Drosophil
c	11	49.4	1.9	1204	17	CNS016E2	AL106628 Drosophil
	12	49.2	1.9	200	13	BJ323623	BJ323623 BJ323623
	13	49.2	1.9	348	9	AU262228	AU262228 AU262228
	14	49.2	1.9	407	13	BJ364861	BJ364861 BJ364861
	15	49.2	1.9	492	13	BJ397639	BJ397639 BJ397639
c	16	49.2	1.9	576	17	CNS035N7	AL228940 Tetraodon
	17	49.2	1.9	878	17	CNS0187R	AL108993 Drosophil
	18	49	1.9	564	10	BE606733	BE606733 WHE0907_E
	19	49	1.9	581	10	BE606185	BE606185 WHE0913_H
	20	49	1.9	707	14	BQ807231	BQ807231 WHE3588_D
	21	49	1.9	724	14	BQ805841	BQ805841 WHE3571_F
c	22	49	1.9	1036	17	CNS00599	AL057797 Drosophil
	23	48.8	1.8	614	13	BJ414131	BJ414131 BJ414131
	24	48.8	1.8	903	17	CNS00I1C	AL074609 Drosophil
	25	48.6	1.8	240	9	AU038043	AU038043 AU038043
	26	48.4	1.8	344	13	BJ326314	BJ326314 BJ326314
	27	48.4	1.8	1007	17	CNS06X9S	AL419462 T3 end of
c	28	48.4	1.8	1101	17	CNS012TP	AL102007 Drosophil
	29	48.2	1.8	496	14	C92059	C92059 C92059 Dict
	30	48.2	1.8	611	13	BJ391725	BJ391725 BJ391725
c	31	48.2	1.8	803	17	B11974	B11974 T24I21-Sp6.
c	32	48.2	1.8	809	17	AQ026987	AQ026987 CIT-HSP-2
c	33	48.2	1.8	997	17	CNS005TE	AL060767 Drosophil
	34	48	1.8	528	17	AZ846659	AZ846659 2M0146O19
	35	48	1.8	628	13	BJ335675	BJ335675 BJ335675
	36	48	1.8	632	13	BJ333944	BJ333944 BJ333944
	37	48	1.8	634	13	BJ397533	BJ397533 BJ397533
	38	47.8	1.8	561	12	BF292807	BF292807 WHE2165_B
	39	47.6	1.8	478	13	BJ335163	BJ335163 BJ335163
	40	47.4	1.8	565	10	BE424450	BE424450 WHE0078_A
	41	47.2	1.8	427	14	C91312	C91312 C91312 Dict
	42	47.2	1.8	476	13	BJ387622	BJ387622 BJ387622
c	43	47.2	1.8	508	17	BH055712	BH055712 RPCI-24-9
	44	47.2	1.8	860	17	CNS007W4	AL050910 Drosophil
	45	47	1.8	518	10	AV980938	AV980938 AV980938

ALIGNMENTS

RESULT 1
CNS0039G/c
LOCUS CNS0039G 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL063921
VERSION AL063921.1 GI:4941778
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial

```
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR08K10"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
BASE COUNT      201 a     64 c    131 g    202 t    503 others
ORIGIN

  Query Match             2.2%;  Score 58.4;  DB 17;  Length 1101;
  Best Local Similarity   13.7%;  Pred. No. 0.0069;
  Matches   71;  Conservative  269;  Mismatches  178;  Indels    2;  Gaps    2;

Qy   1013 AAAATAGAACTTATATTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATA 1072
           ::: :: :: :|:::::: :   :::: |:::::|:|::: ::: | |: | |:::::
Db   1098 RRWGDDTWDRDTRKDDWDWTKWWTWWKDRADDRRWAGDADRWAWDDGAGTWWTATWWWWW 1039

Qy   1073 AGTATTATAATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCA 1132
           : :||::|::: ::: :|| |  :  : :::   : ::|  ::||  :: : |  |:
Db   1038 WWWATWDTWWD-KWWWWATAAKTDTAWTWWRTAWRADWAGRDRGAGKRDRDAATDADGAG 980

Qy   1133 CCATTATGTCTGGATTGGTTTTCCACTCACAACCAATCAATGATAGGCCAAAGCAGGCAT 1192
             :  :::: ::::::|:::::    : |  |  :|| :::::::::   ::|  :: |:
Db    979 RRDGGRKRKDKKDRKDGDDDKKGGKKKKAAKAAKWATKWWDDWDWDKDWKWDGAKDRKAD 920

Qy   1193 GGTGTTGGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCA 1252
           :::|  :::::|: ::|:::|:|  :::::  :::::   | :|:|  | |  ::: : :
Db    919 DDDGAGDKDDDGKGKDADDDTDGTKDDDDKDKWDDWDKAKGTWGDATWAWAATDWWWWGW 860

Qy   1253 AACATCCCAGGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGAG 1312
           |: ::   :   :::  :::: :|  ::: : : |: |: ::: ::| ::    :::::
Db    859 ADADWWTWDAAADDWWADDRWDAWAWKWDDAWAWGARTADRRDWGDRAGKRGGARKRRDR 800

Qy   1313 GAGGAGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTA 1372
           :: :::::   ::::: |  ||  | | :: :::::: : : :: |: |    :: ::::
Db    799 KRADDKRD-AADDRDDAATWTTWTTTTTRDTDDWKWKTDTWTRWAADRTWDRDDDDDRDR 741

Qy   1373 AAATGAATTGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGG 1432
           |   |: : : ::  : :::|: ::|:|::  ::::: : ::  :|   :  |  :::::
Db    740 AGTAGRKWRRTWKRRWKRRDTRWDDADADDTARDDRRRRGDDGADAGKGKKTGRKRRRRD 681

Qy   1433 AACAGCATAAAAGGAATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATA 1492
           :| ::  :|:::|: |::|  :|:  :::: : : :  : :::| |  |: : |::  ::|
Db    680 RATWDRTDAWWADAAWWTTTDTDTDDWDKRDRRRKGARRRRRTTARAAWDWWTWKAWDWA 621

Qy   1493 AAGTAGGCAAAAATGTTTATTTGCCTCCAAGAGAGGGAGA 1532
           :::  : ::::::  :|:|: :  :  :|:|:| : :::
Db    620 KWDWKTRADRWDRWAADTWTDARKADRDWAKARAWRARRD 581


RESULT 2
AU039090
LOCUS       AU039090                592 bp    mRNA    linear   EST 29-MAR-1999
DEFINITION  AU039090 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
            discoideum cDNA clone SSM373, mRNA sequence.
ACCESSION   AU039090
VERSION     AU039090.1  GI:3985843
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 592)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     The Dictyostelium developmental cDNA project: generation and
            analysis of expressed sequence tags from the first-finger stage of
            development
  JOURNAL   DNA Res. 5 (6), 335-340 (1998)
  MEDLINE   99156227
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp
            PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES             Location/Qualifiers
     source          1. .592
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SSM373"
                     /clone_lib="Dictyostelium discoideum SS (H.Urushihara)"
                     /dev_stage="slug"
BASE COUNT      313 a     73 c     26 g    180 t
ORIGIN

  Query Match             2.0%;  Score 52.6;  DB 9;  Length 592;
  Best Local Similarity   52.5%;  Pred. No. 0.14;
  Matches  115;  Conservative    0;  Mismatches  104;  Indels    0;  Gaps    0;

Qy    535 ACAAAATCAATAACAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGAC 594
          |||| | ||| |||||||||| || |||||||| |||| ||| || || ||| |  |
Db    138 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 197

Qy    595 ATGGTCAATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAG 654
          |   |     | | || ||| |  |      |  ||||    | |       || |  |
Db    198 ACAATTTGATAAATTAATTCATTAAATTTAAAAAATAAAATAAAAATAAATAAAAATAAA 257

Qy    655 CAAATGATAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTAC 714
           |||| | ||  |  ||| | || |  | |  |   || |   |  | || ||| |  |
Db    258 AAAATAAAAAAATAAAAAATAAAAAATAAAATGAATAATAATAAAGATAATAAAAAAAAT 317

Qy    715 AATGAAACTTGGTACTCTGCAGATTTGGTATGTGAACAA 753
          ||| | | |    | | | ||  ||    || | |  ||
Db    318 AATAATAATAATAAATTTTCAATTTACACATTTTAGAAA 356


RESULT 3
AU261463
LOCUS       AU261463                230 bp    mRNA    linear   EST 09-MAY-2002
DEFINITION  AU261463 VS Dictyostelium discoideum cDNA clone VSA412 3', mRNA
            sequence.
ACCESSION   AU261463
VERSION     AU261463.1  GI:20518010
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 230)
  AUTHORS   Urushihara,H., Morio,T., Saito,T., Koriki,E., Ochiai,H., Maeda,M.,
            Takeuchi,I., Kohara,Y. and Tanaka,Y.
  TITLE     Population analysis of cDNAs from unicellular and multicellular
            stages of Dictyostelium discoideum
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .230
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
```

```
                     /clone="VSA412"
                     /clone_lib="VS"
                     /sex="mat A"
                     /dev_stage="vegetative"
BASE COUNT      147 a     38 c     12 g     33 t
ORIGIN

  Query Match             1.9%;  Score 51.4;  DB 9;  Length 230;
  Best Local Similarity   55.2%;  Pred. No. 0.25;
  Matches  100;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy    535 ACAAAATCAATAACAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGAC 594
          |||| | ||| ||||||| || || ||||||||||| | ||| || || ||||    |
Db     50 ACAACACCAACAACAACACCAACAACAACAACATCACCAACAACAGCAACAAATACCGTT 109

Qy    595 ATGGTCAATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAG 654
          | | |  ||   | || || || | ||    |  | ||  |   ||    | ||
Db    110 AAGCTTTATATAATTATTTTTTATTTAAATAAAAAGAAACGATTAGAAAAGCAATTTAGA 169

Qy    655 CAAATGATAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTAC 714
           |||  | ||     |||   || |  | |     ||||| | | || ||||||  |  |
Db    170 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 229

Qy    715 A 715
          |
Db    230 A 230


RESULT 4
CNS0182P
LOCUS       CNS0182P               1101 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN37D10 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL108811
VERSION     AL108811.1  GI:5629115
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACN37D10"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
BASE COUNT      274 a    268 c    128 g     73 t    358 others
ORIGIN

  Query Match             1.9%;  Score 51.4;  DB 17;  Length 1101;
  Best Local Similarity   22.4%;  Pred. No. 0.29;
  Matches  122;  Conservative  184;  Mismatches  236;  Indels    3;  Gaps    1;

Qy    400 AATAATACAGTCACAGAACAGGCAATAGAGGATGTATGGCAACTCTTTGAGACCTCAATA 459
          ||:|| | ||:   :|||  |:  :::|||: : ||    ||   ::|::::    || :

Db    536 AAWAAAAAAGWGKARGAAGGGRRGRBVGAGRGDATAATAAAAAAADDTKDRWDARGAAAR 595

Qy    460 AAGCCTTGTGTAAAATTATCCCCATTATGCATTACTATGAGATGCAATAAAAGTGAGACA 519
          ||   :   ::|| :|:|     |  |   |  |  |  || |:  | |||  | |
Db    596 AAAAADAAARWAAKWTWAAAAARAAAAAAAAAAAAAAAAAGTTKTTAAAAAAAGGGWGKT 655

Qy    520 GATAGATGGGGATTGACAAAATCAATAACAACAACAGCATCAACAACATCAACGACAGCA 579
          |    |:|:|||   : :::   :  :| || || |  |: || || |  ||  | |  :
Db    656 GKGKAADGDGGAAAADAWRRGRGRGRRARAARAAAAADAKRAAAAAAAAAAAAAAAAAAD 715

Qy    580 TCAGCAAAAGTAGACATGGTCAATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACA 639
          :  :  |: ::  :  : ||:  |: ::||   : |::|     :|  :|: |: |
Db    716 AGGKRKWKKGDGKDGGKATKTAAWAKGRKDWGTATAWWTWDATWADTWKAATDTDAKRAA 775

Qy    640 GGCTTGGAACAAGAGCAAATGATAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGAC 699
          |  :::|:  |::| ::|  : :|   |  :|   :| | :| | ::  ::::::| |
Db    776 AGRRKRDARKTARDGGRRARTRRRAWAAGGKRA--RAGARRARAARRAADDRDWDAWAA 832

Qy    700 AAGAAAAAAGAGTACAATGAAACTTGGTACTCTGCAGATTTGGTATGTGAACAAGGGAAT 759
          || |||||| | |  ::  ::: :::  : : |  | ::: :  | : ::: ::::|:
Db    833 AAAAAAAAAAAWTTWRDRWWDWDDWDTFWDDDTTAAWWDDARARARRRRRRRRRRRARR 892

Qy    760 AACACTGGTAATGAAAGTAGATGTTACATGAACCACTGTAACACTTCTGTTATCCAAGAG 819
          || :  ::|: ::::|:   :::||  :| ::  : : ::: :     :  ::  :: ::
Db    893 AARRAADDTDTKDRWADATTDKDTTKTWTTDDDDWDKAKRDRWWAAKADGAWKWRDRARD 952

Qy    820 TCTTGTGACAAACATTATTGGGATGCTATTAGATTTAGGTATTGTGCACCTCCAGGTTAT 879
          :     :: : : :::   |::  : :  ::: :|  :::::::::     :  :  || :
Db    953 DWAATAKDDDGWKDKWGGRGKRGGKDKKRWDKGKTGKKDDDDWDKTWTRDWWWWWTRTKW 1012

Qy    880 GCTTTGCTTAGATGTAATGACACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTG 939
          : ::::   :|::   ::|| : |:|: ::: :|: | ||::  :: :| :  ::|:::
Db   1013 DWWWDDGRGRGRWTRKRWGAWWRADAWARDDTDGKDTRTADKDRKRTDTTKRDGDDWRKD 1072

Qy    940 GTGGT 944
          :::::
Db   1073 RKDKK 1077


RESULT 5
CNS00KLH/c
LOCUS       CNS00KLH               1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR17F16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL077724
VERSION     AL077724.1  GI:4957302
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
```

```
          found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR17F16"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
BASE COUNT      173 a    202 c    119 g    459 t    148 others
ORIGIN

  Query Match             1.9%;  Score 50.8;  DB 17;  Length 1101;
  Best Local Similarity   45.5%;  Pred. No. 0.41;
  Matches  136;  Conservative   10;  Mismatches  153;  Indels    0;  Gaps    0;

Qy     449 AGACCTCAATAAAGCCTTGTGTAAAATTATCCCCATTATGCATTACTATGAGATGCAATA 508
           |:|    || ||||        ||||  |:  | || |          |  |||    | |
Db     509 ARAAAAAAAAAAAGAGAAAAAAAAAAAAAAWWACAATAAGAACAAGAGACAAGACAAGAGA 450

Qy     509 AAAGTGAGACAGATAGATGGGGATTGACAAAATCAATAACAACAACAGCATCAACAACAT 568
           |||  ||| ||  :|: ||||   || ||||  || || || || |  |   | || |
Db     449 CAAGAAAGAGAGGGRGRGGGGGGGAGAAAAAAAAAAGAAAAAAAAAAAAAAAAGAAAAAAA 390

Qy     569 CAACGACAGCATCAGCAAAAGTAGACATGGTCAATGAGACTAGTTCTTGTATAGCCCAGG 628
           ||   ||  | || |||||| | | | ||   || ||| ||      | | |    |
Db     389 AAAAAGCAAAAGCAACAAAAGAAAAAAAGGGACATAAGAAAAGAGTAAGCAAAATGGAAA 330

Qy     629 ATAATTGCACAGGCTTGGAACAAGAGCAAATGATAAGCTGTAAATTCAACATGACAGGGT 688
           | ||    |||     | ||||| |  ||   | ||     |||   |  |  | |
Db     329 AAAAAAWAACAAAAAAGCAACAAAAAAAAACAAAAAAAAAAAAAAAAAAAGAAAAAAAWAG 270

Qy     689 TAAAAAGAGACAAGAAAAAAGAGTACAATGAAACTTGGTACTCTGCAGATTTGGTATGT 747
           || ||  |  ||  ||| :||  | | | ||  || |:: | :||   |    :|: |
Db     269 CAAGAATCGCAAAATAAAGWGAWAAAATTCAATTTTTGWWTTTKGCNTTTCAATWAWTT 211


RESULT 6
AU061377
LOCUS       AU061377                244 bp    mRNA    linear   EST 20-MAY-1999
DEFINITION  AU061377 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
            discoideum cDNA clone SLD862, mRNA sequence.
ACCESSION   AU061377
VERSION     AU061377.1  GI:4882481
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 244)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp
            PROJECT = Dictyostelium discoideum cDNA project in Japan.
FEATURES             Location/Qualifiers
     source          1. .244
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SLD862"
                     /clone_lib="Dictyostelium discoideum SL (H.Urushihara)"
                     /dev_stage="slug"
BASE COUNT      127 a     51 c     22 g     43 t      1 others
ORIGIN

  Query Match             1.9%;  Score 50.6;  DB 9;  Length 244;
  Best Local Similarity   60.1%;  Pred. No. 0.38;
  Matches   83;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy     455 CAATAAAGCCTTGTGTAAAATTATCCCCATTATGCATTACTATGAGATGCAATAAAAGTG 514
           |||||    ||| | | | || | | | || ||| |  || |  | |  ||  ||||||
Db      99 CAATATGAACTTTTATTAGATCAACTCTATGATGAACAACAACAACAACCACAAAAAGTT 158

Qy     515 AGACAGATAGATGGGGATTGACAAAATCAATAACAACAACAGCATCAACAACATCAACGA 574
           | | | | |     |  |||| | ||| |||| ||||| || || ||||| ||||
Db     159 TCAAATAAACAACAACAGCAACAACANCAACAACATCAACAACAACAGCAACAACAACAT 218

Qy     575 CAGCATCAGCAAAAGTAG 592
           || || || ||||   ||
Db     219 CAACAACAACAAACACAG 236


RESULT 7
CNS0100X
LOCUS       CNS0100X                1101 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN03G04 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL098379
VERSION     AL098379.1  GI:5609990
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACN03G04"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
BASE COUNT      195 a    108 c    131 g    161 t    506 others
ORIGIN

  Query Match             1.9%;  Score 50.6;  DB 17;  Length 1101;
  Best Local Similarity   16.9%;  Pred. No. 0.45;
  Matches  111;  Conservative  268;  Mismatches  273;  Indels    3;  Gaps    2;

Qy     817 GAGTCTTGTGACAAACATTATTGGGATGCTATTAGATTTAGGTATTGTGCACCTCCAGGT 876
           | |  :|:|:  : :  :|:|:|:|: :|: : :|:: |:|:  | |||:: :  :   : :
Db     424 GKGAKWTDTRTDDTWKTDTDTWGKADTDWWTDTWKTTWTWTWTTTTGWKKWDTKTKTKWK 483

Qy     877 TATGCTTTGCTTAGATGTAATGACACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAG 936
           :::| |:|: :::  |::: ::    ::|: | :  :: :|:|::  |  :|::| |: |
Db     484 DDDGTTWTDTWKDWKTDKDTDKTWTDVWADNAGWWTDKDKTKAKKKRTKTDTKKTKTRTG 543

Qy     937 GTGGTGGTCTCTTCATGCACAAGGATGATGGAGACACAGACTTCTACTTGGTTTGGCTTT 996
           :::|:: :  |: :   : :|:: :||  :: | | ::::  : :| |:  :|:| |||
Db     544 TKKKTDRGMRKTDTWGWGDWRADRTDGARADRAANAKDKD-ADGKADTDATKTDGWGTTT 602

Qy     997 AATGGAACTAGAGCAGAAAATAGAACTTATATTTACTGGCATGGTAGGGATAATAGGACT 1056
           |:  :: :|::  :::::::: :    ::: : : ::| ::: |  :: : :: :   :
```

```
Db    603 KADTTDDTDADWWDWKDWDWKDADTKAGRKRGWGWGKKGGWKKTTKTKKKKTDKTKTGTD 662

Qy   1057 ATAATTAGTTTAAATAAGTATTATAATCTAACAATGAAATGTAGAAGACCAGGAAATAAG 1116
          :: :::::|:  ::  : : ::::     : : :   :  :| :|:|  |  |:|||| 
Db    663 TKWGWKDKWTDKKRDGKRGDGDRKDKGTGGDGDSKAKGDRAWAWRAKATRAAAADATAAA 722

Qy   1117 ACAGTTTTACCAGTCACCATTATGTCTGGATTGGTTTTCCACTCACAACCAATCAATGAT 1176
            |: | :|    |      || |:| |   ||  |::|  | : : ::   |: : :::
Db    723 TAAKATAWATKTTTKKTTTTTTTKTTTTTTTTTTTKKTTTARKADRDDDDKAKRDGDKDG 782

Qy   1177 AGGCCAAAGCAGGCATGGTGTTGGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTG 1236
          : :  : :  ||: ::::|:::| ::  | :: |  ::      ::  :   |:| : :|
Db    783 DWDDADKDWKAGRDRDDDTRKWGWKDRAGKKKKAKKKDHKTAWWWADDTGTAWATWAKTA 842

Qy   1237 AAGCAGACCATTGTCAAACATCCCAGGTATACTGGAACTAACAATACTGATAAAATCAAT 1296
            |:  :|   ::::  :         :|   :   : :   : : :  :: ::     : :
Db    843 TAKTGKAKATKDBKRGRGGGGRGDGRGAGRRKRARGRKRGRRRKDKRKRKKRTKKAGDGD 902

Qy   1297 TTGACGGCTCCTGGAGGAGGAGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGA 1356
          |    ::    :: :|:  |:::  :: ::||:  |:  | ::|: ::   :    :::
Db    903 TGAGTDKAGDKKKADGDGWGWKRKWDSKTWKTTMTGTDDTTTWKGRNRRGRGRAGAGRRR 962

Qy   1357 GAGTTCCTCTAC--TGTAAAATGAATTGGTTTCTAAATTGGGTAGAAGATAGGAATACAG 1414
          |:: :  :      :::|:|: :: :| :::: :: :|: |   | :|: ::| |:: |:
Db    963 GRKGKAGKGAGGKWWKWAWADABWTWTWDWKDWKWKWTDWGWGKGGWGDGRKGGAKDGAK 1022

Qy   1415 CTAACCAGAAGCCAAAGGAACAGCATAAAAGGAATTACGTGCCATGTCATATTAG 1469
            :   :|::   :|:: :: |  |:|::|::::: ::::: |: : ::||:::
Db   1023 RNRKNRDGMRYGNRAWKWMDDAATAKARWADDDWDADYDDDADADYWKWKATWMD 1077
```

```
RESULT 8
CNS005IJ/c
LOCUS       CNS005IJ                991 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR12O10 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL059544
VERSION     AL059544.1  GI:4943182
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 991)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .991
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR12O10"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
```

```
BASE COUNT      158 a    133 c    158 g    480 t     62 others
ORIGIN

  Query Match             1.9%;  Score 50.2;  DB 17;  Length 991;
  Best Local Similarity   41.6%;  Pred. No. 0.55;
  Matches  157;  Conservative   21;  Mismatches  199;  Indels    0;  Gaps    0;

Qy    455 CAATAAAGCCTTGTGTAAAATTATCCCCATTATGCATTACTATGAGATGCAATAAAAGTG 514
          ||| |||       | ||||  |     |  |   |  |  |  | |   || ||||
Db    484 CAAAAAAAAAAACGGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAMAC 425

Qy    515 AGACAGATAGATGGGGATTGACAAAATCAATAACAACAACAGCATCAACAACATCAACGA 574
          :|| | | | |     |   |:|:||  || || || || |  |  || || |  ||  |
Db    424 MGAAAAAAAAAAAAAAAAAAAMARAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 365

Qy    575 CAGCATCAGCAAAAGTAGACATGGTCAATGAGACTAGTTCTTGTATAGCCCAGGATAATT 634
          |  |  |  ||||  | | |     ||  | |:     :    | |   :|  | ||
Db    364 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAMAGAACMAAAAAMAAAAMAAAAAAAA 305

Qy    635 GCACAGGCTTGGAACAAGAGCAAATGATAAGCTGTAAATTCAACATGACAGGGTTAAAAA 694
            | | |     || || :  ||:  | || |   |||   || |  | |  :  |||||
Db    304 AAAAAAGAAAAAAARAAAMAAAAWAAAAAAACAAAAAAAAAAAAAAAAAAATRGAAAAAA 245

Qy    695 GAGACAAGAAAAAAGAGTACAATGAAACTTGGTACTCTGCAGATTTGGTATGTGAACAAG 754
          |::     |||||| |  | ||  ||||    |||  |  | |     :|    | | :|
Db    244 GWRGACCCAAAAAAAAAAAAAAAAAAAACMAAWTACMATAAAAAAAAAAWAAAAWAGCTWG 185

Qy    755 GGAATAACACTGGTAATGAAAGTAGATGTTACATGAACCACTGTAACACTTCTGTTATCC 814
            |   |  | ||| | ||  | |:  :| :  || | :| | | | ||  :| :||
Db    184 CTCAACTTAACGTTAACGCAAAAAAAWHAWAWWCMAAACTMTMTTAAACAAMTAWTAAAW 125

Qy    815 AAGAGTCTTGTGACAAA 831
            :    ||||| |:|||
Db    124 CWATAACTTGTAAMAAA 108
```

```
RESULT 9
CNS0106X
LOCUS       CNS0106X                1101 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN03K20 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL098595
VERSION     AL098595.1  GI:5610206
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACN03K20"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : T7"
BASE COUNT      258 a    107 c     60 g    175 t    501 others
```

```
ORIGIN

  Query Match             1.9%;  Score 50;  DB 17;  Length 1101;
  Best Local Similarity   20.0%;  Pred. No. 0.62;
  Matches   97;  Conservative  187;  Mismatches  201;  Indels    1;  Gaps    1;

Qy    1023 TTATATTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTTAAATAAGTATTATA- 1081
           :| |:|::|   :  | ::|: :||:||  :  |:|:    :| :: |:  :|:| |||
Db     594 WTTTMTWKAKGTKTGTTDKAKTKATMATGTKTATRTWTAATKTAKRTADTKDTRTAATAT 653

Qy    1082 ATCTAACAATGAAATGTAGAAGACCAGGAAATAAGACAGTTTTACCAGTCACCATTATGT 1141
           |: |   |::: :  ::|  ::     :|  :||:| |:   :    :: :    :|: :
Db     654 AKATKKKARKDKDTAKDAWKWDKATKAKAKKKAAKAKAKAAAKTATGKDARKAKAKAKAK 713

Qy    1142 CTGGATTGGTTTTCCACTCACAACCAATCAATGATAGGCCAAAGCAGGCATGGTGTTGGT 1201
           |::|: : ::|     | | ||      :  :|:|:   :: : ::   ::::: : :
Db     714 ATKRAKARAWDTATWTATAADAAADKGAKDKAKAKAKADARDKRRWDAKKDKRKRAKAKA 773

Qy    1202 TTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGAAGCAGACCATTGTCAAACATCCCA 1261
           :  :: :|:|::  | :|:  |: |:|:: :|:|  : |::   :::::  |: ::
Db     774 AKAADDADADAKDAKAADAKAKADAADADDGRGDGKKKRAKDRKKKKKKDKAWGDKKKAK 833

Qy    1262 GGTATACTGGAACTAACAATACTGATAAAATCAATTTGACGGCTCCTGGAGGAGGAGATC 1321
           :::| |  : :|  :| | |  :  |::: |   |   ::|  : :   : | :|::|
Db     834 KDKAAAKAKADAAGDAKAKRRRAGDKDKAKADAKAAKKAKAKDDDDAKAATKAKKATKAK 893

Qy    1322 CGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAGAGTTCCTCTACTGTAAAATGAATT 1381
           ::|  ::   ::  : : :| | : :: :: :::::::  : :  : : ::|:: |::
Db     894 KDKAKAKKKKKKKDKDAKAKAKAKDAKDKDDDKDDKKDAKADKKKWKKDRAKDKAKK 953

Qy    1382 GGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTAACCAGAAGCCAAAGGAACAGCATA 1441
           ::::: :| :::::::: :|:|:|:  |:| |: :|   |:||:  |||:  | |  |:|
Db     954 DDDKKDKAKDKDKDKKDADAKAWAKAKADADAKAKAKAKAKAADDDAAADAKAKAADAKA 1013

Qy    1442 AAAGGAATTACGTGCCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAGGCA 1501
           |  :|:: |  :   : :  : : :|:|  :|  |: :| | : :  |:: :::|:: :
Db    1014 KAKAKADDAAKAKAKADAKAADADAKAKAAKRAAKAKADADAAKAKAKADDKDKDAKKDD 1073

Qy    1502 AAAATG 1507
           |: |::
Db    1074 ADKAKK 1079


RESULT 10
CNS00EVL
LOCUS       CNS00EVL                1101 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069706
VERSION     AL069706.1  GI:4949849
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR29B23"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
BASE COUNT      419 a     91 c     60 g     299 t    232 others
ORIGIN

  Query Match             1.9%;  Score 49.4;  DB 17;  Length 1101;
  Best Local Similarity   33.3%;  Pred. No. 0.86;
  Matches  158;  Conservative   68;  Mismatches  249;  Indels    0;  Gaps    0;

Qy     314 ATACTTGGGGAACAACTCAGTGCCTACCAGATAATGGTGATTATTCAGAAGTGGCCCTTA 373
           ::| |:    || || :   |   |:  | |:|::     || :: :  : |     : :
Db     470 WWAATWTWWAAAWAAAWTTATWAATWAAAAAAWWWWWATTTTTWWWWWTWATTWTTTWAW 529

Qy     374 ATGTTACAGAAAGCTTTGATGCCTGGAATAATACAGTCACAGAACAGGCAATAGAGGATG 433
           :| | : | |||     :       |  |::|||| | | | |    |   ::|| |  |
Db     530 WTWTAWTAAAAAAAAAWATAATTTAAAWWAATAWATTAAWAATTTAWAAWWTATATTAAT 589

Qy     434 TATGGCAACTCTTTGAGACCTCAATAAAGCCTTGTGTAAAATTATCCCCATTATGCATTA 493
           :     |  |  || | |    || | |   || : |||||| |    | | |    | :
Db     590 WTATAAATWTWATTAATATAAAAAAATATTTTTTWATAAAATTTTTAATAATTTAATTAW 649

Qy     494 CTATGAGATGCAATAAAAGTGAGACAGATAGATGGGGATTGACAAAATCAATAACAACAA 553
           |||  | ||  : :|  : : |   | |||  |    ||  | ||||  || || |:  :
Db     650 TTATTAAATAAWTWATTWWWTAATTAAATAATTTWAAATAWAAAAAAAAAAAAAAAWATW 709

Qy     554 CAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGACATGGTCAATGAGACTAGTT 613
           |  || |  :| |       | :     |  ||:| |: : ||  |  ::     || ||
Db     710 AAWAATWATAWATAAWTTAAAAWAATAAAAWAAWAATWAWWATAATATWWATATATAWTT 769

Qy     614 CTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAGCAAATGATAAGCTGTAAAT 673
           |  ::      :  |||: |  |      :  |:  | |  :|   ||:|  : :|:||
Db     770 WTAWWWATWWAWWWTATAWATAWAATAWAAWAWAWATAAATAWATAWATWAAAWAWAWAT 829

Qy     674 TCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTACAATGAAACTTGGTACTCTG 733
           :| :  | |    ||:|:| | |  : || | : :| :| ||: |||  |   : | :
Db     830 AWWATWATATAWWAATAWAWAAAAAATWTAATATWAATWATAAWAAAAAATAWAWTTWWT 889

Qy     734 CAGATTTGGTATGTGAACAAGGGAATAACACTGGTAATGAAAGTAGATGTTACAT 788
           |::  :|| | || |:   :| |: |     ||  ||   : :: |:| ||
Db     890 WTTTTWWAWWATATAAAWAWATAWAAAWAAAAAAAAAAAAAATAAWAWWWTWATAT 944


RESULT 11
CNS016E2/c
LOCUS       CNS016E2                1204 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN15A12 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL106628
VERSION     AL106628.1  GI:5622852
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1204)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
```

collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.

```
FEATURES             Location/Qualifiers
     source          1. .1204
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACN15A12"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : T7"
BASE COUNT      298 a    172 c    106 g    316 t    312 others
ORIGIN
```

```
  Query Match             1.9%;  Score 49.4;  DB 17;  Length 1204;
  Best Local Similarity   23.3%;  Pred. No. 0.87;
  Matches  112;  Conservative  155;  Mismatches  212;  Indels    2;  Gaps    1;

Qy     998 ATGGAACTAGAGCAGAAAATAGAACTTATATTTACTGGCATGGTAGGGATAATAGGACTA 1057
           |: :||  | |  | |||| |:||  :| | ::      |  : |:: | :| |    :|
Db    1169 AKWKAAAAAAAAAAAAAAAAAAKAAAADAAAAKKKAAAAAAAAKAADDAAADAAAAAKAKA 1110

Qy    1058 TAATTAGTTTAAATAAGTATTATAATCTAACAATGAAATGTAGAAGACCAGGAAATAAGA 1117
             |: |    |:| : ::|: | ||  :||  ::    ::::::  |  |::   :|  :
Db    1109 AKAKAAAAAAADAAWKKKAWAAAAAAKWAAAKWKATKKDDDDKDKAAADAKKKKKKAKAD 1050

Qy    1118 CAGTTTTACCAGTCACCATTATGTCTGGATTGGTTTTCCACTCACAACCAATCAATGATA 1177
            ::  :::   :  |  :::|::  ::: : ::::::  : | | :|  ||:  :::|:
Db    1049 ADKAADKDKDKKAKADAWDDADDAAKKKKKAKKKDKKKDDATDADWAAKAAKKKDKKAWK 990

Qy    1178 GGCCAAAGCAGGCATGGTGTTGGTTTGGAGGAAAATGGAAGGATGCAATAAAAGAGGTGA 1237
           ::       : ::::::::::::::: :::  |:::   |::'|| ||||  ::::
Db     989 KKKDKKKAKKAKAWKKKDKKDKKDKKKKKKKDKKAKKKKKAAAKDAAAAAAAAAKKKDKK 930

Qy    1238 AGCAGACCATTGTCAAACATCCCAGGTATACTGGAACTAACAATACTGATAAAATCAATT 1297
            : |:|   :::: ||   :    :: : |  ::|| :   ||:  :: ::||:: ||::
Db     929 KDAAKAAAKKKKKDAAKKKKAKKKDDAWAADADKAAAKKKAAAKKKDKKKDAADKKAAKD 870

Qy    1298 TGACGGCTCCTGGAGGAGGAGATCCGGAAGTTACCTTCATGTGGACAAATTGCAGAGGAG 1357
             :: :  :  :::| :  ::  :      |  ::  :| |:   |  | | | ||
Db     869 AAKKKKDKAAKAADKDAAKKWDKKKKAGGAARAAAKKGKKGGGRKGGGAGGGRAAAAGAT 810

Qy    1358 AGTTCCTCTACTGTAAAATGAATTGGTTTCTAAATTGGGTAGAAGATAGGAATACAGCTA 1417
           : |:  | |  |:  :| |   ::  |:  :| |:: ::|: ||:: :|::|   |:  :
Db     809 WTTWAKTDTTKTKAGDAKTTTTKKTTTWATKAGAKWAKKTRAAAKDARGKDAATTAKAAR 750

Qy    1418 ACCAGA--AGCCAAAGGAACAGCATAAAAGGAATTACGTGCCATGTCATATTAGACAAAT 1475
           |  | |   |  |:|||||  : | |:|' |:|| :|  |:  |  |   :  ||: :| :
Db     749 AAGAAATKGGGAARAGGAAGGKDAGARAKGKAAAKAGATKGAAAATACMWAAAGRAWAGW 690

Qy    1476 A 1476
           |
Db     689 A 689
```

```
RESULT 12
BJ323623
LOCUS       BJ323623                200 bp    mRNA    linear   EST 05-MAR-2002
DEFINITION  BJ323623 Dictyostelium discoideum cDNA library, AF Dictyostelium
            discoideum cDNA clone ddalln11 5', mRNA sequence.
ACCESSION   BJ323623
VERSION     BJ323623.1  GI:19153753
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 200)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the aggregation
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .200
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddalln11"
                     /clone_lib="Dictyostelium discoideum cDNA library, AF"
                     /sex="mat A"
                     /dev_stage="Aggregation stage"
BASE COUNT      120 a     45 c      7 g     26 t      2 others
ORIGIN
```

```
  Query Match             1.9%;  Score 49.2;  DB 13;  Length 200;
  Best Local Similarity   63.6%;  Pred. No. 0.79;
  Matches   75;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy     471 AAAATTATCCCCATTATGCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGG 530
           | ||| || |  || | | |    |||    | ||| | ||   |   | ||| |||
Db      13 ATAATAATTCAAATCAAGAAAATTTATCTTTTCCAAAAGAAAATAATAATATAAATGTAA 72

Qy     531 ATTGACAAAATCAATAACAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAA 588
           || | ||| | ||| |||||||||| || |||||||| |||| ||| || || ||| ||
Db      73 ATAGTCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 130
```

```
RESULT 13
AU262228
LOCUS       AU262228                348 bp    mRNA    linear   EST 09-MAY-2002
DEFINITION  AU262228 VS Dictyostelium discoideum cDNA clone VSB761 3', mRNA
            sequence.
ACCESSION   AU262228
VERSION     AU262228.1  GI:20518775
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 348)
  AUTHORS   Urushihara,H., Morio,T., Saito,T., Koriki,E., Ochiai,H., Maeda,M.,
            Takeuchi,I., Kohara,Y. and Tanaka,Y.
  TITLE     Population analysis of cDNAs from unicellular and multicellular
            stages of Dictyostelium discoideum
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .348
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="VSB761"
                     /clone_lib="VS"
                     /sex="mat A"
                     /dev_stage="vegetative"
BASE COUNT      196 a     71 c     19 g     61 t      1 others
ORIGIN
```

```
  Query Match             1.9%;  Score 49.2;  DB 9;  Length 348;
  Best Local Similarity   54.4%;  Pred. No. 0.84;
  Matches   99;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;
```

```
Qy     535 ACAAAATCAATAACAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAAGTAGAC 594
           |||| | ||| ||||||||||| || ||||||||||||| ||| |  || |||||  | ||
Db     164 ACAACAACAACAACAACAACAACAACAACATCAACAACAACTACAACAAAAAAAAAC 223

Qy     595 ATGGTCAATGAGACTAGTTCTTGTATAGCCCAGGATAATTGCACAGGCTTGGAACAAGAG 654
                   || ||  |  | ||||    |   |       |   |      |  ||||
Db     224 CAAAAATGAAGGAATATGTAGTTTATAATAAATATTTTCAATATTAGTCACTATAAAGAA 283

Qy     655 CAAATGATAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAGACAAGAAAAAAGAGTAC 714
            |||  | ||     |||   || |  | |     ||||| | | || ||||||  |  |
Db     284 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 343

Qy     715 AA 716
           ||
Db     344 AA 345


RESULT 14
BJ364861
LOCUS       BJ364861                407 bp    mRNA    linear   EST 08-MAR-2002
DEFINITION  BJ364861 Dictyostelium discoideum cDNA library, CF Dictyostelium
            discoideum cDNA clone ddc33c11 5', mRNA sequence.
ACCESSION   BJ364861
VERSION     BJ364861.1  GI:19274161
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 407)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the culmination
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .407
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddc33c11"
                     /clone_lib="Dictyostelium discoideum cDNA library, CF"
                     /sex="mat A"
                     /dev_stage="Culmination stage"
BASE COUNT      203 a     64 c     29 g    108 t      3 others
ORIGIN

  Query Match             1.9%;  Score 49.2;  DB 13;  Length 407;
  Best Local Similarity   63.6%;  Pred. No. 0.86;
  Matches   75;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy     471 AAAATTATCCCCATTATGCATTACTATGAGATGCAATAAAAGTGAGACAGATAGATGGGG 530
           |||   || |  ||| |   | |   | |    ||||.|||   | |||  || |||  |
Db      92 AAAGAAATTCAAATTTTATTTCAAATTTATTAACAATTAAAAATAAACAATTAAATGTTG 151

Qy     531 ATTGACAAAATCAATAACAACAACAGCATCAACAACATCAACGACAGCATCAGCAAAA 588
           ||| |  ||||||| ||||||||||    |||||||| |||| ||| || || ||| |
Db     152 ATTCATCAAATCAACAACAACAACAATTACAACAACAACAACAACAACAACAACA 209


RESULT 15
BJ397639
LOCUS       BJ397639                492 bp    mRNA    linear   EST 08-MAR-2002
DEFINITION  BJ397639 Dictyostelium discoideum cDNA library, SF Dictyostelium
            discoideum cDNA clone dds47d02 5', mRNA sequence.
ACCESSION   BJ397639
VERSION     BJ397639.1  GI:19308725
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 492)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the slug stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .492
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="dds47d02"
                     /clone_lib="Dictyostelium discoideum cDNA library, SF"
                     /sex="mat A"
                     /dev_stage="Slug stage"
BASE COUNT      210 a    107 c     51 g    122 t      2 others
ORIGIN

  Query Match             1.9%;  Score 49.2;  DB 13;  Length 492;
  Best Local Similarity   60.4%;  Pred. No. 0.88;
  Matches   81;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy     455 CAATAAAGCCTTGTGTAAAATTATCCCCATTATGCATTACTATGAGATGCAATAAAAGTG 514
           ||  |||   |||   ||||  |||| ||    | | |||||  |  | ||  | ||
Db     301 CATCAAAATTTTGAAGAAAAAGATCCACAAATAGAAATACTACCATCTCCACAACAACAA 360

Qy     515 AGACAGATAGATGGGGATTGACAAAATCAATAACAACAACAGCATCAACAACATCAACGA 574
             |||   | |     |   |||| | ||| ||||||||| || |||||||| |||| |
Db     361 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAA 420

Qy     575 CAGCATCAGCAAAA 588
           || || || ||| |
Db     421 CAACAACAACAACA 434


Search completed: November 16, 2002, 13:07:29
Job time : 2470.32 secs
```